Structural Studies to Enable Drug Discovery

A Dissertation
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

WILLIAM MICHAEL MCCUE

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BARRY CRAIG FINZEL, ADVISOR

August 2021

Acknowledgements

I would like to begin by thanking my advisor, Dr. Barry C. Finzel, for his constant guidance and support. I appreciate you constantly challenging me and making me think critically about projects even when I felt they were reaching a dead end. Thank you for showing me the beauty that is structural biology and for giving me an appreciation for how easy we structural biologists have it today. Thank you to the lab mates that I have had along my graduate journey, but especially Dr. Kimberly Maize. Thank you for teaching me all the laboratory techniques that I needed to accomplish what I did and for mentoring me so that I could lead the next round of graduate students.

I would like to thank the Department of Medicinal Chemistry here at the University of Minnesota for allowing me to further my education and giving me the tools and opportunities to become the scientist I am today. I would like to thank my cohort of graduate students that I began this journey with 6 years ago, especially Dr. Anand Divakaran and Ellie Mews for their emotional support and always having a cold one ready after a long week.

My work presented in this thesis is the result of collaborative efforts of which I was just a part. The CD44-THIQ-Monosaccharide conjugate project was a collaboration with the Georg and McCarthy laboratories and benefitted from the wonderful synthetic work of Dr. Soma Maitra. The modified-HA as inhibitors of CD44 was a collaboration with the Huang laboratory at Michigan State University. The Caspase project brought together a large team including the laboratories of Dr. Karen Ashe and Dr. Michael Walters. The work by Dr. Kathryn Nelson was instrumental in advancing the biochemistry involved in inhibitor design.

I detail 12 new protein structures in this dissertation which would have not been possible without access to APS 17-ID: IMCA-CAT and the knowledge of their support staff, especially Anne Mulichak and Eric Zoelner.

I would like to thank my committee members: Dr. Gunda Georg, Dr. James McCarthy and Dr. Elizabeth Ambrose. Gunda and Jim, thank you for your patience and wisdom as we attempted to develop an inhibitor of a rather difficult target, CD44. Thank you, Elizabeth, for helping me in use of computational methods to rationalize negative results.

I would like to thank my truly loving and supportive friends and family for believing in me and coming along with me through this journey. Thank you to my friends who have supported me while not understanding a word of what I do. Thank you to my sister (and her amazing family) for their love and support. To my in-laws, thank you for welcoming me into your family with such open arms. To my parents, thank you for instilling in me a strong work ethic and for pushing me to be my best self. Thank you for highlighting the importance of an education and using my talents for the betterment of society. Finally, I would like to thank my amazing wife, the first Dr. McCue. Thank you for being there with me through this journey; I could not have done it without you, and I love you more than you will ever know.

Dedication

To my parents and my wife.

Abstract

To better understand molecular interactions upon small-molecule binding, three different targets (MDH1, CD44, and Cas-3) for intervention in disease states were studied using biophysical and structural biology techniques. The detailed findings will hopefully guide future inhibitor design to provide better treatment options for individuals suffering from cancer and tauopathies.

Malate dehydrogenase I (MDH1) is a cytosolic NAD-dependent oxidoreductase that has recently garnered attention for supplementing lactate dehydrogenase (LDH) in supplying adequate cytosolic NAD levels necessary during rapid cellular growth in cancer progression. The first ever human MDH1 crystal structure is described in detail and differences in the active site loop relative to MDH2, the mitochondrial isoform, could potentiate selective inhibitor design.

CD44 is of the family of hyaladherins that is overexpressed in a variety of different cancers. Antagonism of CD44-HA interactions should block downstream signaling and therefore inhibit cancer development, progression, and metastasis. Three different approaches are discussed in the dissertation. The first uses structure-based drug design to link a small-molecule to a saccharide portion of HA. Second, SPR was used to quantify binding affinity of a series of modified-HA molecules. Finally, biophysical techniques were used to confirm binding of verbascoside to the HABD of CD44.

The Caspases are of the family of endoproteases involved in apoptosis and pyroptosis pathways to maintain homeostasis. Caspase-3 was used as a surrogate for Caspase-2 (only Caspase to cleave tau at Asp314) to generate co-crystal complexes with a series of pentapeptides. Described structures show the binding modality of the pentapeptides and highlight conformational differences in the L4 loop.

Table of Contents

List of Tables

List of Figures

List of Schemes

List of Equations

1. Introduction

1.1 Macromolecule X-Ray Crystallography in Drug Discovery

1.1.1 Receptor Hypothesis of Drug Action

Pharmacology as a scientific discipline was born in the mid-19th century with the first pharmacology department being established in Buchheim in 1847 to meet the need of understanding how therapeutic drugs and poisons produced their effects.[1] The inadequacy of understanding therapeutic drugs was summed up in Oliver Wendell Holmes' comment in 1860: "… if the whole materia medica as now used could be sunk to the bottom of the sea, it would be all the better for mankind – and the worse for the fishes".[2] In 1865, Kekulé discovered the structure of the benzene ring and the now familiar 2-dimensional representations of the structure of organic molecules began to be defined which paved the way for lock-and-key relationships between drugs and their receptors.[1] The receptor theory hypothesis was first put forward by John Newport Langley in 1905 where he used the term "receptive substance" as distinct from the term "contractile substance".[3] The term "receptive substance" was used to explain the actions of nicotine and curare on skeletal muscle based on his experiments on salivary secretion in dogs.[4] A. V. Hill, a student working in Langley's laboratory, expressed the receptor idea quantitatively in terms of a bimolecular reaction following the law of mass action which later became the well-known Hill-Langmuir equation.[5,6] Paul Ehrlich, the father of immunology, through his interest in the immunology and chemotherapy of infectious diseases, came to the idea of receptors.[7] Bacterial toxins interact with nutrient-capturing structures of cells, "sidechains" which he later called "receptors", thus starving the cells which respond by making more of these "sidechains" and releasing them into circulation where they combine with the toxin rendering them harmless.[8] He later postulated that bacterial sidechains and sidechains of the host differ which led him to the discovery of Salvarsan in 1909, the first effective treatment for syphilis.[9,10] Receptor theory was published by Alfred Joseph Clark in his 1933 book, "The Mode of Action of Drugs on Cells" where

he calculates for potent drug-like acetylcholine, the number of molecules required to effect a cellular response is insufficient to cover the cell and therefore concluded there must be a specific receptor interaction at the cell surface that mediate the cellular response.[11]

1.1.2 History of X-Ray Crystallography

The history of X-rays traces back to Germany in 1895, where William Röntgen first generated X-rays by means of a discharge from a large induction coil passed through a Hittorf's vacuum tube or through a well-exhausted Crookes' or Lenard's tube.[12,13] The discharge was found to penetrate black cardboard and caused an observed fluorescence still visible at two meters.[12,13] In his 1895 Science paper, Röntgen describes this fluorescent discharge as "X-rays (as I will call the rays, for the sake of brevity)" and qualified the transparency of several different metals and media for which he was awarded the Nobel Prize in Physics 1901.[12,13] It was 17 years later that Max von Laue, with the help of his two assistants, postulated through his work on a blue crystal of copper sulfate pentahydrate that the diffraction of X-rays might be related to the interatomic distances.[14] Laue was unable to provide an explanation of the diffraction images, but a father-son duo, Sir William Henry Bragg and Sir William Lawrence Bragg, were empirically able to relate diffraction patterns to crystal structure.[15] William Lawrence Bragg developed a mathematical formula that became known as Bragg's Law, $n\lambda = 2d\sin\theta$, describing the relationship between the angle of refraction, the wavelength of radiation, and interplanar spacing in the crystal lattice.[15] Constructive interference of a diffracted wave will be observed when the distance traveled through the crystal is an integer (n) multiple of the wavelength (λ) and that the path length through the crystal is twice the interplanar space (d) in the crystal multiplied by the sine of the angle of incidence (θ). The formula has withheld the test of time and is paramount to the interpretation of crystallographic data.[16–19] The elder Bragg, William Henry Bragg, went on to construct the first X-ray spectrometer.[17] One of the first structures solved by the duo was that of diamond.[20] The work of Max von Laue, Sir William Henry Bragg and Sir William Lawrence Bragg was immediately

recognized by the scientific community with Laue receiving a Nobel Prize in Physics in 1914 and both of the Braggs in 1915 (Sir William Lawrence Bragg received the award at 25 and he is still the youngest person to win the Nobel Prize in Physics).[14] The Nobel Prize in Physics has been awarded 23 times for crystallography or for essential technology in how it is used and of the 23 awards, 12 have been awarded directly for achievement in macromolecule crystallography.[14]

1.1.3 X-Ray Diffraction for Determination of Biological Crystals

The first diffraction image for a biological structure was of pepsin protease generated by J. Desmond Bernal and Dorothy Hodgkin (previously Crowfoot) in 1934.[21] The diffraction was of poor quality as the crystals were mounted on glass fibers and exposed to air as was customary for small molecule crystals.[21] They therefore decided that crystals needed to be hydrated and therefore sealing them in capillaries with a drop of mother liquor could efficiently protect them from desiccation. Of the images collected, Bernal noted that: "… the [X-ray] pictures yielded by protein crystals were of exceptional perfection. They showed large unit cells with great wealth of reflections […] found even at comparatively high angles corresponding to such low spacings as 2 Å. This indicated that not only were the molecules of the proteins substantially identical in shape and size, but also that they had identical and regular internal structures right down to atomic dimensions".[22] The structure of the pepsin protease in this form was not determined until 1990 by Hodgkin's former student, Sir Tom Blundell.[22] Hodgkin's impact on biomolecular crystallography cannot be understated: she collected the first diffraction images of insulin[23], determined the structure and stereochemistry of sterols[14], showed that penicillin contained a β-lactam ring[24], and solved the structure of the first organometallic compound, vitamin B12, work for which she was awarded the 1964 Nobel Prize in Chemistry.[25] The first proteins to have their complete structures solved by X-ray crystallography were sperm whale myoglobin, in 1958, and horse hemoglobin, in 1960.[26,27] The structures were solved by Sir John Kendrew and Max Perutz (a pioneer in the methodology of protein crystallography especially isomorphous replacement)[28], respectively, and

only extend to 6 Å resolution. In the two decades following the solving of the myoglobin and hemoglobin structures, several protein structures with clinical relevance were published including dogfish lactate dehydrogenase[29], bacterial thermolysin[30], bovine pancreatic phospholipase A_2[31], and bacterial dihydrofolate reductase[32]. The Protein Data Bank was established in 1973 and contained nine protein structures, six of which were of interest in drug design, showing the impetus of the early stages of structural biology.[14,33] The bovine pancreatic trypsin inhibitor, aprotinin, which is administered to control bleeding during major surgery, was one the first structures deposited into the PDB.[34,35]

1.1.4 Application of Structure-Based Drug Design Using Crystallography

Wootton's work on compounds designed to fit a site in hemoglobin merged the receptor theory, a drug molecule interacting with a specific receptor of a complementary structure, with structural biology.[36] In the study, the known three-dimensional coordinates of hemoglobin with 2,3-diphosphoglycerate (DPG) were used to generate a molecular model to design compounds which should bind to the deoxy conformation and thereby promote oxygen liberation.[36] The designed compounds, using structural data, were found to promote oxygen liberation and thus mark the first known compounds designed using structure-based design.[36] The compounds ended up proving unsuccessful as no anti-sickling compound ever came to the market. The concept of structure-based drug design continued to grow in importance leading to the rational drug design cycle being elaborated: a molecule can be designed and optimized using structural data of the macromolecule target followed by co-crystallization of the molecule with the target of interest and then the cycle can be repeated.[37] This cycle is still used in the drug discovery process with additional sophistications including experimental confirmation of binding, ADMET evaluation, and biological application.[38]

The first structure that contributed to the design of a marketed drug was bovine pancreatic carboxypeptidase A. The work done by the Lipscomb lab revealed that the preferred substrate of

the zinc metalloprotease is the C-terminal residue of a polypeptide chain with a preference for aromatic or branched aliphatic sidechains.[39] The active site is marked by a deep pocket-like topography with the presence of an arginine residue to interact with the free carboxylate, and a large mostly hydrophobic pocket to accommodate the P1' sidechain.[39] The arginine when interacting with a substrate carboxylate induces a large conformational change in a nearby tyrosine that essentially closes the active site.[39] This structure-function relationship was used by scientists at Squibb to build a model of the active site of angiotensin-converting enzyme (ACE) whose function is the cleavage of two C-terminal amino acids from angiotensin I to transform into angiotensin II which causes immediate increase in blood pressure.[40] The scientist at Squibb used the model of carboxypeptidase and the known ACE substrate to design potent inhibitors of ACE eventually leading to captopril which came to market in 1981 as the first drug using structure-based design.[41,42]

1.2 Targets for Cancer Intervention

Cancer is the second leading cause of death worldwide behind cardiovascular diseases. This year, an estimated 1,806,590 new cases of cancer will be diagnosed in the United States and 606,520 people will die from the disease.[43] The most common cancers are breast cancer, lung and bronchus cancer, prostate cancer, and colon and rectum cancer.[43] Approximately 39.5% of men and women will be diagnosed with cancer at some point during their lifetimes.[43] Estimated expenditures for cancer care in the United States in 2018 were $150.8 billion and in future years, costs are likely to increase as the population ages and more people develop cancers.[43] There is a need for novel cancer therapeutics to help patients living with cancer while decreasing the cost of cancer care. Cancer progression, metastasis, and motility has been extensively studied and characterized. Normal cells are tightly regulated by hundreds of genes that intricately control the process of cell division.[44] Carcinogenesis is started through a variety of environmental and inherited characteristics that can cause damage to DNA thus causing mutations in somatic cells.[44] These cells can undergo

a series of events including activation of growth-promoting oncogenes, impaired apoptosis, and inactivation of tumor suppressor genes which can lead to an altered gene product with abnormal structural and regulatory proteins which is a hallmark of a malignant tumor.[44] Cancer cells are characterized by large, variably shaped nuclei, disorganized arrangement, varying size and shape, and loss of normal features.[44] The role of CD44 and malate dehydrogenase I (MDH1) in cancer progression and metastasis will be discussed in detail in the subsequent two sections.

1.3 CD44

1.3.1 Role of the Extracellular Matrix in Cancer

The extracellular matrix (ECM) is the non-cellular space present within all tissues and organs that provides the scaffolding for the cellular constituents and initiates biochemical signals that play a role in morphogenesis, differentiation, and homeostasis.[45] Every tissue has a unique ECM composition that arises from tissue development and constant remodeling to maintain extracellular homeostasis.[45] The ECM is composed of two main classes of macromolecules: fibrous proteins and proteoglycans (PGs) which are made up of glycosaminoglycan (GAG) chains linked to a protein core (with the exception of hyaluronan).[46,47] The ECM is highly charged and hydrated and contributes to compression strength and flexibility.[48] The biochemical properties of the ECM allow for the control of signaling molecules interaction with their cognate receptors.[49] ECM remodeling is tightly regulated through controlling the expression/activity of ECM enzymes.[50,51] The activities of the remodeling enzymes can be deregulated with senescence and certain disease states such as cancer.[52] The deregulation of the ECM dynamics leads to disorganization as the amount, composition, and/or topography becomes abnormal.[52] One of the main contributors to the altered activities of ECM remodeling enzymes are cancer-associated fibroblasts (CAFs).[53] Deregulation of the ECM can potentiate oncogenic effects on signaling pathways, malignant cell transformations, and increased tissue stiffness which promotes tumor cell invasion and progression.[52] Proper regulation of the ECM remodeling enzymes could decrease tumor

progression and metastasis to different tissues. Targeting abnormal ECM as an effective cancer therapeutic requires an understanding of how the ECM is organized and maintained and how it becomes deregulated in cancer.

1.3.2 Hyaluronan: Signaling Molecule of the ECM

Hyaluronan (HA) is a major component of the ECM where it has been shown to promote cell motility, adhesion, and proliferation and thus plays a critical role in wound repair, inflammation, and metastasis.[54,55] Hyaluronan is a GAG synthesized by a family of three transmembrane glycosyltransferases, hyaluronan synthetase 1-3 (HAS1-3).[56,57] HAS1-3 are plasma membrane bound proteins that can alternatively conjugate (β-1,4)-glucuronic acid (Glc) and (β-1,4)-*N*-acetylglucosamine (NaG) to form the polymeric molecule that is extruded into the ECM.[56,57] 4-methylumbelliferone blocks HA synthesis through the inhibition of HASs and was found to cause tumor suppression in breast cancer cell lines.[58] Inhibition of HASs is not a focus of this thesis but highlights the importance of hyaluronan in the extracellular matrix and cancer. Under physiological conditions, HA exists as a high-molecular weight polymer $>10^6$ Da where it regulates a variety of cell behaviors, including cell adhesion, motility, and growth.[59–63] HA levels are tightly regulated to maintain homeostasis and approximately 30% of circulating HA is degraded and replaced by newly formed HA every 24 hours.[64] Degradation of HA occurs by endocytic uptake which is triggered by binding to CD44 and/or lymphatic vessel endothelial receptor-1 (LYVE-1).[65,66] Hyaluronidases (HYAL) degrade HA to oligosaccharides which are then digested by glucuronidase and hexosaminidase to yield more supply of Glc and NaG, respectively, that can be reused by HASs to form new HA molecules.[67,68] Degradation of HA is increased by reactive oxygen species and pathological conditions such as cancer.[69,70] HA fragments, $<5 \times 10^5$ Da, generated during inflammation/tumorigenesis/tissue injury can act as a signaling molecule by alerting the injured host that HA homeostasis has been affected.[69–71] Degradation of HMW-HA to HA fragments has

been linked to cell differentiation[72], angiogenesis[73], tumorigenesis[74], and resistance to cancer chemotherapy[75].

1.3.3 Hyaladherin Family of Proteins

The diverse roles of HA, described in the previous section, are mediated through hyaladherins, or hyaluronan-binding proteins (HABPs).[76,77] Of the HABPs, CD44 and RHAMM (receptor for hyaluronic-acid-mediated motility) are signal-transducing receptors that influence cell proliferation, survival and motility.[76,77] Cellular responses resulting from interactions of HA with CD44 and RHAMM involve tyrosine kinases, protein kinase C, focal adhesion kinase (FAK), phosphatidylinonsitol 3-kinase (PI3K), mitogen-activated protein kinase (MAPK), nuclear factor-κB and Ras, as well as cytoskeletal components.[76,78–80] The hyaladherins can be localized on cell membranes (such as LYVE-1 and CD44) or can be in the ECM or cytoplasm (such as TSG-6 and RHAMM).[77] As discussed previously, various lengths of HA can trigger diverse biological responses because of recognition with a variety of hyaladherins.[71] Many of the HABPs contain a common structural domain termed a Link module that is involved in hyaluronan binding.[77] The Link module was first characterized in the link protein isolated from cartillage.[81] The Link module can be subdivided into three groups based on the size of the binding domain: Type A consists of a single Link module, e.g., TSG-6; Type B consists of a Link module with N- and C-terminal extensions, e.g., CD44; and Type C consists of a Link module pair, e.g., Link protein.[82] The consensus fold of the Link module is two α-helices and two triple-stranded anti-parallel β-sheets. A BX_7B motif, where B corresponds to an arginine or lysine residue and X is any non-acidic amino acid, has been implicated in binding of hyaluronan in hyaladherins regardless of the presence of the Link module.[83] Selectively targeting a hyaluronan binding domain of a hyaladherin poses a unique challenge due to the commonality in structure and sequence.

1.3.4 HABD of CD44

CD44 is a membrane-associated glycoprotein of the family of hyaladherin with a Type B Link module. The protein consists of three domains: an extracellular hyaluronan binding domain (HABD), a transmembrane domain, and an intracellular domain. 14 isoforms of CD44 exist that vary in the transmembrane domain. The intracellular domain or cytoplasmic domain is involved with signaling cascades including the PI3K and MAPK pathways.[84] The 14 different isoforms of CD44 include a consensus functional HABD of approximately 150 residues that are conserved.[84] The structure of the HABD and the interaction it makes with oligosaccharides of HA has been studied using structural biology. The murine CD44 (*m*CD44) and human CD44 (*h*CD44) HABD, recombinantly expressed from *Escherichia coli*, have been crystallized.[85,86] Structural efforts have employed the murine isoform (86% sequence identity with the human protein) of the CD44 HABD domain as the crystal packing allows for soaking of small molecules into binding sites of the protein (Figure 1.1). A structure of *m*CD44 with an oligosaccharide of HA (PDBid: 2JCR), shows that the HABD contains an extended binding groove (800 $Å^2$) joined by β2-α1, α2-β3, and β4-β5 loops that can accommodate binding of HA.[85] Ten residues of the Link module (Arg45, Tyr46, Cys81, Arg82, Tyr83, Ile100, Ala102, Ala103, His105 and Tyr109) make hydrogen bonding interactions with the oligosaccharide (Figure 1.2). The HABD has no well-formed or deep pockets that would serve as attractive binding sites for small-molecule inhibitors and is known to undergo small but important conformational changes upon binding HA.[85] The protein-polysaccharide complex resembles protein-protein interactions that are difficult to effectively disrupt with small-molecules.[87] Previous biophysical studies have quantified the K_D of HA-8 (an oligosaccharide in length of 8 saccharide units) using isothermal titration calorimetry (ITC) to be 125 μM and using surface plasmon resonance (SPR) to be 16.7 μM.[85,88] CD44 responds differently to HA dependent on the length of HA: polymeric-HA promotes structural integrity[89,90] while HA fragments (<500 kDa) trigger

signaling cascades resulting in stimulation of the innate immune response[91], recruitment of immune cells[90] and tumor growth[74,75,92,93].

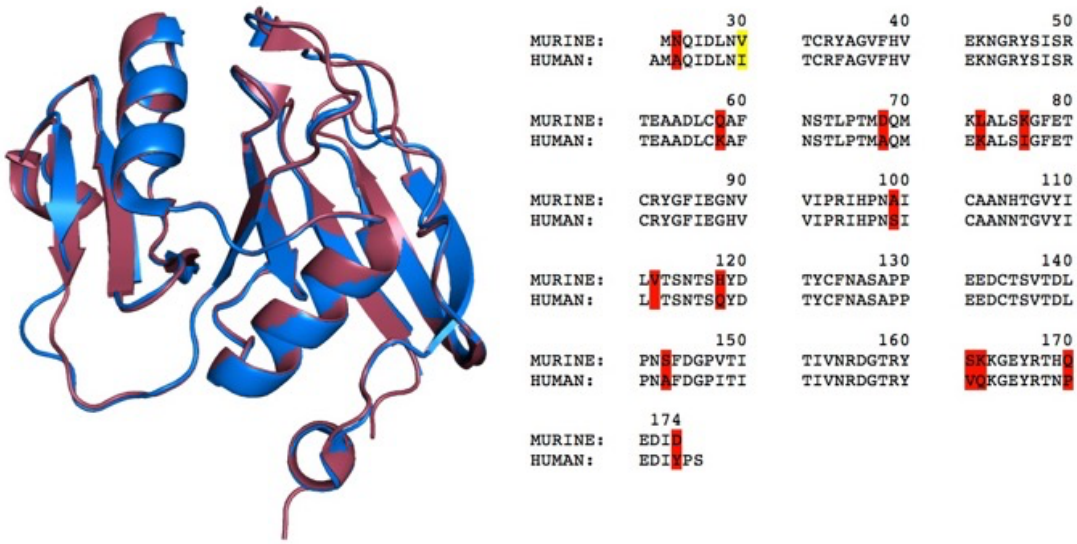


Figure 1.1 Comparison of *h*CD44 and *m*CD44

Overlay of *h*CD44 (PDBid: 4PZ4) in magenta and *m*CD44 (PDBid: 5BZC) in blue. Amino acid differences between the two isoforms are highlighted in red and conserved changes in yellow.

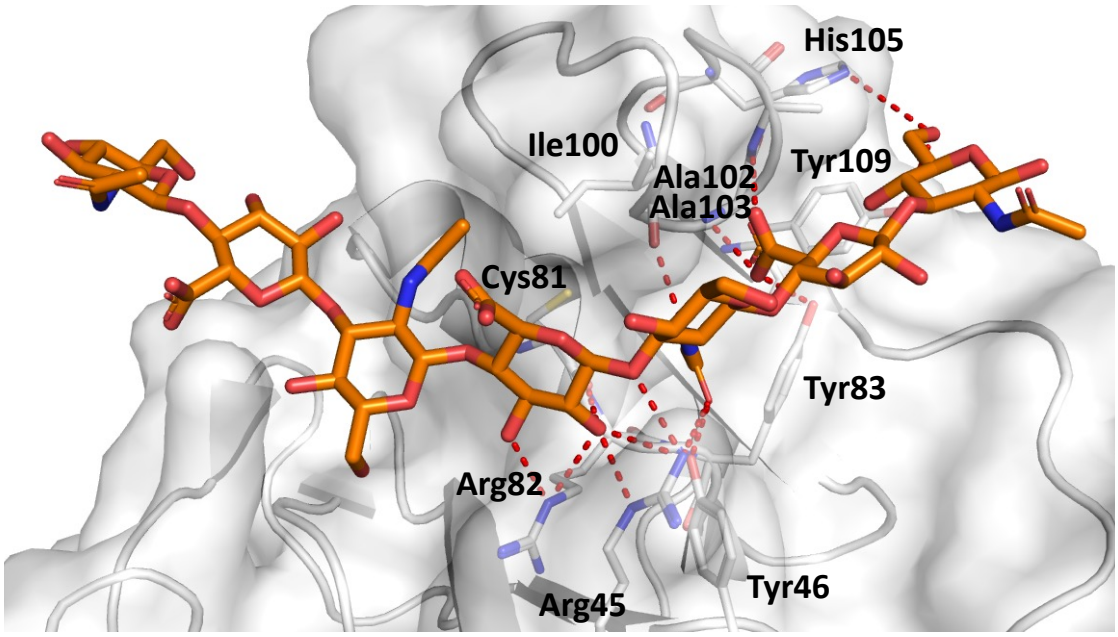


Figure 1.2 Interaction of HA within Hyaluronan Binding Groove (PDBid: 2JCR)

1.3.5 Antagonism of CD44-HA Interactions

Selective antagonism of CD44-HA interaction could be a useful treatment to slow progression and metastasis in certain types of cancer.[63] Several different approaches have been taken to interrupt HA-CD44 interactions. Monoclonal antibodies that block CD44 binding to HA were found to reduce tissue swelling and inflammation by inhibiting immune cell infiltration.[94,95] In clinical trials, anti-CD44 antibodies, against overexpressed CD44 variants, conjugated with a cytotoxic molecule can selectively deliver the drug to cells expressing the specific variants and thus can inhibit CD44 interactions/signaling and can cause apoptosis.[96–100] Unfortunately, these conjugated antibodies caused toxicity leading to phase I clinical trials being halted. Other than antibodies, some small-molecules have been used to antagonize the CD44-HA interaction. Short HA-oligosaccharides (6-18 saccharide units) were able to inhibit downstream cell survival and proliferation pathways through competitive inhibition of CD44 interactions with full-length HA.[92,101–103] Work in the Finzel laboratory has led to the discovery of an inducible small-molecule subsite on CD44 HABD that preferential binds molecules containing a tetrahydroisoquinoline

(THIQ) pharmacophore.[88] These THIQ molecules were able to antagonize CD44-HA interactions with low millimolar IC_{50} values.[88] Information on the structure-based design of the THIQ-series of molecules is discussed in greater extent in Chapter 4. Recently, verbascoside was found to inhibit CD44-dimerization and computational studies postulated that the natural product was able to interact with the HA-binding groove of CD44 (discussed in more detail in Chapter 5).[104]

1.4 Malate Dehydrogenase I (MDH1)

1.4.1 The Warburg Effect

Normal differentiated cells metabolize glucose to carbon dioxide by oxidation of glycolytic pyruvate in the mitochondrial tricarboxylic acid (TCA) cycle. This series of reactions produces NADH from NAD which can then be used to drive oxidative phosphorylation to maximize production of adenosine triphosphate (ATP), the "energy currency" of the cell. ATP is considered "energy currency" as the hydrolysis of ATP to adenosine diphosphate (ADP) yields Gibbs-free energy of -7.3 kcal/mol which provides the energy for many essential processes including intracellular signaling, DNA and RNA synthesis, purinergic signaling, active transport, and muscle contraction. In the absence of oxygen, cells undergo anaerobic respiration where pyruvate is reduced to lactate yielding the oxidation of one NADH molecule to NAD^+ for a less efficient generation of ATP (2 moles of ATP/mole of glucose in anaerobic respiration compared to ~36 moles of ATP/mole glucose in oxidative phosphorylation). In 1924, Otto Warburg discovered that unlike normal tissues, cancer cells metabolize glucose into lactate even in the presence of sufficient oxygen concentrations to support oxidative phosphorylation. The process is now referred to as the Warburg effect or aerobic glycolysis. Converting glucose into lactate instead of into pyruvate subverts the TCA (cycle) and allows for more efficient synthesis of biomass through anabolism necessary for rapid cellular growth and cellular division, hallmarks of cancer progression.

1.4.2 NAD(P)H-Dependent Oxidoreductases

The NAD(P)H-dependent dehydrogenases are a family of enzymes that use either nicotinamide adenine dinucleotide (NAD) or nicotinamide adenine dinucleotide phosphate (NADP) as cofactors in enzymatic activity and are found in many of the steps involved in oxidative phosphorylation, anerobic glycolysis, and aerobic glycolysis.[105] The NAD(P)H-dependent oxidoreductases are able to oxidize their substrates by transferring a hydride group to NAD^+ and many oxidoreductases are able to complete the reaction in the opposite direction (transferring a hydride group from NADH) to reduce the substrate.[105] The nicotinamide ring of NADH/NAD^+ is directly involved in the transfer of electrons during NADH oxidation.[106] The C4 position of the nicotinamide ring is able to act as the acceptor/donor of the proton during enzymatic activity.[106] The presence of the phosphate group on NADPH modifies the structures of the cofactor and allows for difference in enzyme specificity for a given cofactor.[105] Typically NADH/NAD^+ are used in catabolic pathways while NADPH/$NADP^+$ enzymes are used for anabolism.[107,108] NADH/NAD^+ ratio depends upon the availability and redox state of external electron acceptors while NADPH/$NADP^+$ pool is maintained in a more reduced state to efficiently provide power for biosynthesis.[109,110]

Many NAD(P)H-dependent oxidoreductases exist, but the majority contain a common structural motif that is the nucleotide binding domain of the enzyme. The most common fold of this structural motif is the Rossmann fold but a variety of less common motifs exists that can also bind the cofactor.[111] The Rossmann fold, named after Michael Rossmann who discovered the fold in lactate dehydrogenase, is a conserved motif present in other NAD(P)H-binding enzymes.[112,113] The Rossmann fold consists of two parallel β-strands separated by an α-helix with a tight loop between the first β-strand and α-helix which establishes direct contact with the cofactor (Figure 1.3).[113] The loop contains a consensus sequence of G-X-G-X-X-G/A (X is any amino acid).[114,115] The glycine-rich loop hydrogen bonds with the phosphate portion of the NAD(P)H/NAD(P). Computational

analysis revealed an invariably water molecule position that bridges the phosphate to the glycine-rich loop.[116] The oxidoreductases typically contain two instances of the Rossmann fold: one interacts with the adenine moiety and the other with the nicotinamide ring. The Rossmann fold can be extended with additional β-strands, usually parallel strands but antiparallel strands are also observed, to form a larger β-sheet.[117] The catalytic domain, which coordinates the substrate and provides the residues for the redox reaction to occur, is more variable amongst the NAD(P)H-oxidoreductases.[105]

Malate dehydrogenase (MDH) is an NAD-dependent oxidoreductase that reversibly catalyzes the reaction of malate to oxaloacetic acid using NADH as a cofactor.[118] Humans express two forms of the enzyme: MDH1 that is localized in the cytosol and MDH2 that is localized in the mitochondria and is involved in the TCA cycle.[118] MDH1 participates in the malate/aspartate shuttle that shuttles reducing equivalents instead of NAD/NADH across the mitochondrial membrane.[118] Kinetics have shown that the enzymatic activity is an ordered reaction with NAD/NADH binding first followed by the binding of the substrate, malate/oxaloacetic acid.[119]

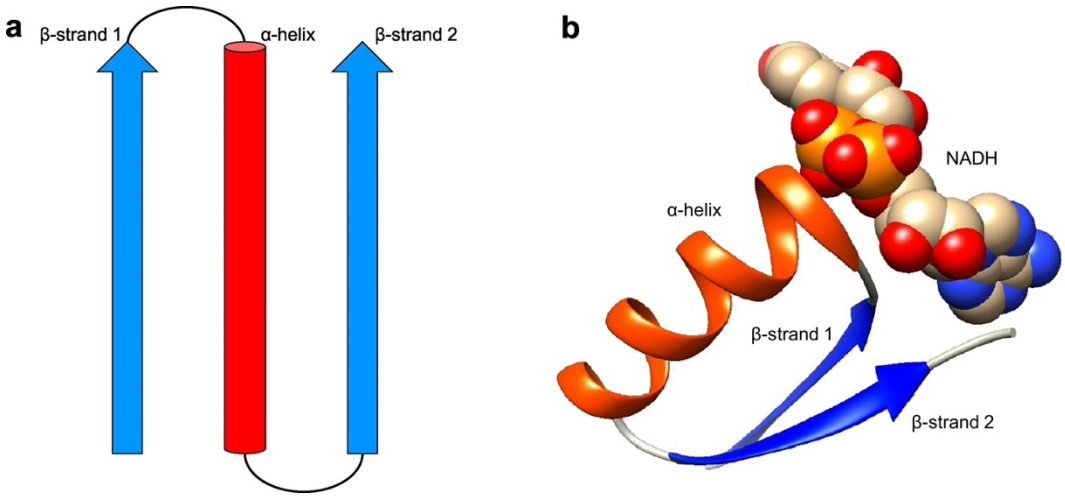


Figure 1.3 Rossmann Fold Structural Motif

a. Diagram of the secondary structure topology of the core Rossmann fold. b. NADH contacts the tight loop formed between β-strand 1 and the α-helix. Figure 1.3 was used with permission from: Vidal, L. S.; Kelly, C. L.; Mordaka, P. M. Heap, J. T. Review of NAD(P)H-dependent oxidoreductase: Properties, engineering and application. *Biochimica et Biophysica Acta – Proteins and Proteomics* **2018**, *1866* (2) 327-347.

1.4.3 Malate Dehydrogenase I (MDH1) Works with LDHA in The Regeneration of Cytosolic NADH

Cancer cells are known to exhibit increased glucose consumption while uncoupling glycolysis from the TCA cycle and shunting carbons from glucose into the biosynthesis of macromolecules to support the rapid growth and proliferation characteristic of cancer progression.[120] To meet the need of increased rate of glycolysis, a constant supply of NAD is required.[120] The cytosolic NAD serves as an electron acceptor for glyceraldehyde-3-phosphate dehydrogenase (GADPH) which catalyzes the conversion of glyceraldehyde-3-phosphate to glycerate-1,3-biphosphate, the sixth step in glycolysis.[120,121] The cytosolic NAD/NADH pool is separate from the mitochondrial pool of NAD/NADH that is involved in the electron transport chain.[120] The regeneration of cytosolic NAD from NADH has been attributed to lactate dehydrogenase (LDH) that converts pyruvate to lactate.[120,122] The shunting of glucose carbons into

biomass means that less carbon atoms can make it through glycolysis meaning lower concentrations of pyruvate and therefore other enzymatic interactions are required to maintain cytosolic NAD levels.[123,124] Recent efforts by the Kelekar laboratory in the Department of Laboratory Medicine and Pathology at the University of Minnesota has led to the discovery that *h*MDH1 complements LDH in replenishing cytosolic NAD levels for enhanced glycolysis during cancer progression.[125] Using a previously established isogenic model, deuterium flow was traced from glucose isotopomer, [4-^{2}H]-glucose, to cytosolic NADH and metabolites from NADH-dependent dehydrogenase activity.[124,125] The largest concentrations of deuterium were traced to lactate, aspartate, and fumarate where the latter two originate from malate.[125] The assay points to the fact that MDH can support LDH in supplying the cytosolic level of NAD. Jurkat cells knocked-out for MDH1 proliferate at a slower rate and consume less glucose than their wild-type counterparts.[125,126] Use of The Cancer Genome Atlas (TCGA) database and cBioPortal, showed that MDH1 is amplified in almost 11% of one squamous cell lung carcinoma dataset.[127,128]

1.4.4 Structural Comparison of the Malate Dehydrogenases from a Variety of Species

While the human cytosolic human malate dehydrogenase I (*h*MDH1) has not been structurally characterized, several other structures of homologous proteins have been solved and deposited into the Protein Data Bank (PDB).[129] Structural data has been described for the mitochondrial isoform from *E. coli* (PDBid: 5KKA[130]; 3HHP[131]; 2CMD[132]), plants (PDBid: 1SMK[133]), and mammals (*sus scrofa*, PDBid: 1MLD[134]) and cytosolic isoform from several species including bacteria (PDBid: 1BMD[135]; 4TVO[136]), plants (5NUE[137]), and mammals (*sus scrofa*, PDBid: 5MDH[138]). The human isoforms of malate dehydrogenase share only 26% sequence identity but share common structural features. The MDHs are stable as dimers with each monomer containing two structurally and functionally distinct domains.[118,132] Located near the N-terminus, the NAD-binding domain contains the Rossmann fold motif with two additional β-strands to form a parallel β-sheet.[132,139] The substrate binding site and amino acids that are necessary for catalysis

are located in the other C-terminal domain.[132] The active site is the cleft formed by these two domains.[132] Molecular binding of the cofactor is stabilized through a series of hydrogen bonds with the amino acids in the active site (Gly13, Gln14, Ile15, Ser88, Asn130, His186, and Ser240; Residue numbers for *ss*MDH1 (PDBid: 5MDH)), (Figure 1.4). The dimer interface consists of mainly interacting α-helices.[140] Formation of the enzyme-cofactor-substrate ternary complex causes a protein conformational change in which a loop closes over the active site to screen the substrate and catalytic residues from the solvent.[141,142] The active site contains a histidine/aspartate pair that provides proton transfer and binding of the substrate carboxylate is anchored by the guanidinium portion of an arginine.[143] MDHs have been shown to distinguish between their substrates and other structurally similar hydroxy-acid substrates in the TCA cycle: MDH is six orders of magnitude less efficient using pyruvate and lactate than its natural substrate which can be attributed to the mobile loop switch.[138] This highlights the molecular recognition of the MDHs which could be exploited to selectively inhibit MDH over other oxidoreductases.

*h*MDH1 as a target for the development of cancer therapeutics (potentially co-administered with an LDH inhibitor) has been potentiated by the work done by Dr. Ameeta Kelekar's laboratory.[125] To enable structure-guided drug design of *h*MDH1, structural data is required. Structural data of the human isoform will allow for the understanding of the binding modality of the substrates and cofactor and enable the design of molecules which can perturb the enzymatic activity as cancer therapeutics. Comparison of the structures of the mitochondrial and cytosolic protein can be used to gain selectivity for MDH1 and thus decrease off-target interactions.

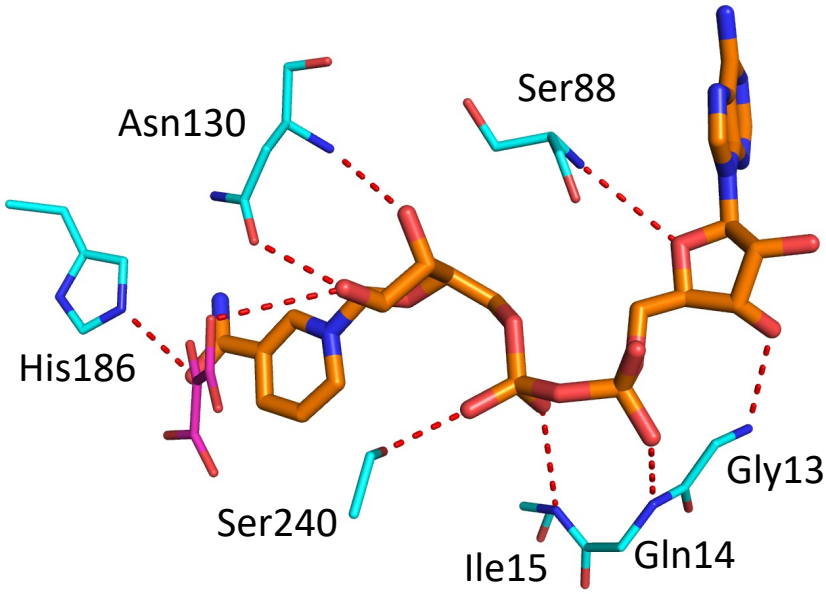


Figure 1.4 Hydrogen Bonding Network of NADH in MDH1

Figure 1.4 was used with permission from: McCue, W.; Finzel, B. C. Structural characterization of the human cytosolic malate dehydrogenase I. *ACS Omega*. Manuscript Submitted.

1.5 Caspase-3

1.5.1 Tauopathies

Tauopathies is a broad term that refers to diseases that are plagued by irregular deposition of microtubule-associated-protein tau (commonly referred to as tau).[144–146] Tau levels are elevated in a variety of disease states including Alzheimer's disease, progressive supranuclear palsy, corticobasal syndrome, frontotemporal dementias, and chronic traumatic encephalopathy, to name a few.[147] Alzheimer's disease (AD) is the most common tauopathy, and according to the Alzheimer's Association more than 6 million Americans are living with AD (1 in 9 people age 65 or older).[148] Total payments for all individuals with Alzheimer's is estimated at $355 billion and is expected to rise to $1.1 trillion in the year 2050.[148] Short term memory loss (forgetting conversations or recent events) is the first sign of AD.[147] As the disease progresses, severe memory impairment and the inability to carry out everyday tasks will be observed in AD patients.[147] There

are some medications that can slow the progression of AD symptoms, but there is currently no available therapeutic that can cure AD.[147]

1.5.2 Caspase Family of Proteins

Apoptosis or programmed cell death plays a role in the development and homeostasis of all multicellular organisms.[149,150] The term 'apoptosis' is credited to John Kerr, Andrew Wyllie, and Alastair Currie in 1972 where they used it to describe a mode of death associated with fragmentation of genomic DNA.[151] Along with fragmentation of genomic DNA, apoptosis is characterized by cytoplasmic and chromatin condensation, nuclear pyknosis, cell rounding, membrane blebbing, cytoskeletal collapse and the formation of membrane-bound apoptotic bodies.[152] Studies involving *Caenorhabditis elegans* (*C. elegans*) found that 131 out of 1090 somatic cells undergo programmed cell death during normal development leading to the discovery of cell lineage-dependent programmed cell death.[153,154] The central components of the programmed cell death machinery in *C. elegans* are three cell death abnormal proteins (commonly referred to as CED). In 1993, the cysteine protease encoded by the *ced-3* gene was found to be structurally similar to the human interleukin-1b-converting enzyme (later renamed Caspase-1).[155–157] Overexpression of Caspase-1 was found to be sufficient to induce apoptosis in mammalian cells.[158] To date, 11 human caspases (Caspase-1 through Caspase-10, and Caspase-14) have been discovered and classified.

Caspases are of the family of endoprotease enzymes that are involved in maintaining homeostasis, regulating cell death and inflammation, and developing normal organ functions.[159] Caspases nucleophilically cleave enzymes at cysteine and aspartate residues which is where they get their name Caspases (**C**ysteine **as**partic **p**rote**ases**) and were first discovered by Nobel Laureate Robert Horvitz.[157] The Caspase family of proteins share a number of characteristics that separate them from other proteases. Caspases are synthesized as inactive zymogens that are activated by proteolytic cleavage to remove the prodomain and separate the p20 large subunit and p10 small

unit (Figure 1.5).[160] Activation of the caspases can follow an intrinsic or extrinsic pathway based on the origin of the signal.[161] The catalytic dyad is present on the p20 subunit and consists of an active site Cys285, which is a part of the conserved Q-A-C-X-G pentapeptide sequence, and His237 (numbering for Caspase-1).[162] Caspases recognize at least four contiguous amino acids and cleave after the C-terminal residue, usually an aspartic acid residue. The pentapeptide that covalently binds to the catalytic cysteine during cleavage is referred to as P1 and the amino acids extending from the N-terminus are called P2, P3, etc. Caspases can be classified by two methods. First, Caspases can be classified on the basis of their known major functions: pro-apoptotic (Caspase-2, -3, -6, -7, -8, -9, 10) and pro-inflammatory (Caspase-1, -4, -5).[160] The pro-apoptotic Caspases, as the name implies, are mainly involved in mediating cell death signaling transduction, while the pro-inflammatory Caspases regulate cytokine maturation during inflammation.[160] Second, the Caspases can be classified by the length of their prodomains which corresponds to their positioning within the apoptotic signaling pathway: initiator Caspases (Caspase-1, -2, -4, -5, -8, -9, -10) and executioner Caspases (Caspase-3, -6, -7).[160] The initiator Caspases contain a long prodomain that contains either the death effector domain (DED) or the caspases recruitment domain (CARD).[160] This subgroup of Caspases are involved with interacting with the upstream signaling molecules within the apoptosis cascade. Executioner Caspases, on the other hand, contain a short prodomain and are typically processed and activated by upstream initiator Caspases before performing execution steps of apoptosis by cleaving multiple cellular substrates.[160]

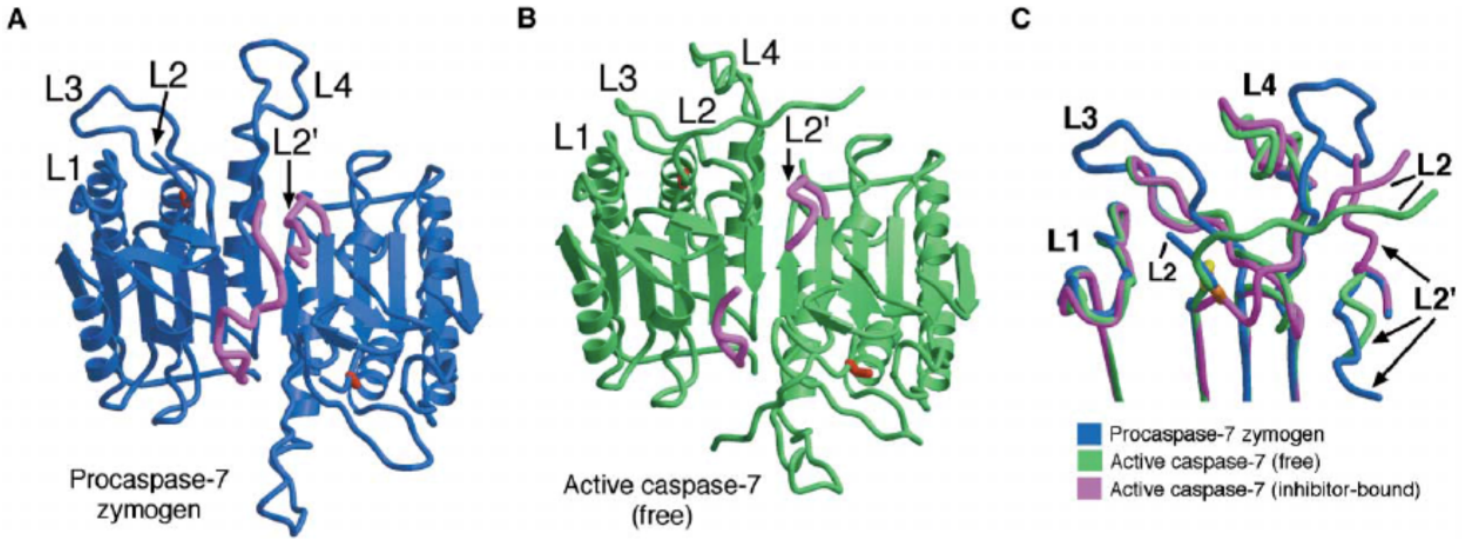


Figure 1.5 Caspase Activation

(A) Structure of a procaspase-7 zymogen (PDB code 1K86). Compared to that of the inhibitor-bound caspase-7, the conformation of the active site loops does not support substrate binding or catalysis. The L2′ loop, locked in a closed conformation by covalent linkage, is occluded from adopting its productive and open conformation. (B) Structure of an active and free caspase-7 (PDB code 1K88). The active site loops are still flexible. Despite an interdomain cleavage, the L2′ loop still exists in the closed conformation, indicating an induced-fit mechanism for binding to inhibitors/substrates. (C) Comparison of the conformation of the active site loops. Compared to the procaspase-7 zymogen or the free caspase-7, the L2′ loop is flipped 180° in the inhibitor-bound caspase-7 to stabilize loops L2 and L4. Figure 1.5 was used with permission from: Shi, Y. Mechanisms of Caspase Activation and Inhibition during Apoptosis. *Molecular Cell* **2002**, *9* (3), 459-470. License Number: 512159030700

1.5.3 Role of Cas-2 in Neurodegenerative Diseases

Microtubule-associated-protein (MAP) tau (commonly referred to as tau or tau protein) functions primarily in stabilizing microtubules (MT) especially in the axons of neurons.[163] There are six major isoforms that arise from alternative splicing from a single encoding gene.[163] The C-terminal portion contains a MT-binding domain that is composed of multiple instances of a highly conserved tubulin-binding domain.[164] The rest of the protein is characterized by a basic proline-rich region and a highly acidic N-terminus region referred to as the 'projection domain'.[164] Studies have substantiated a model where the tubulin-binding domains bind to specific pockets on β-tubulin at the inner surface of the MTs, the positively charged proline-rich regions are bound to the negatively charged MT surface, and the highly acidic residues near the N-terminus protrude from the MT surface due to charge-charge repulsion.[165,166] β-tubulin pockets of adjacent protofilaments can be occupied by different repeats of the same MT-binding domain of tau thus crosslinking multiple MTs to further contribute to MT stabilization.[165] Binding affinity of tau for MTs is regulated by serine/threonine-directed phosphorylation.[167] Kinase-mediated phosphorylation detaches tau from MT while dephosphorylation reaction by phosphatases restores MT-binding ability of tau.[167] Phosphorylation of tau leads to the increase of cytosolic unbound tau where it becomes more susceptible to undergo misfolding and thus more prone to aggregation referred to as pretangles.[168–170] Pretangles can transition to form paired helical filaments (PHFs) and then self-assemble to form neurofibrillary tangles (NFTs).[171]

Neurodegenerative disorders, where tau no longer binds to the MTs and instead it becomes sequestered into NFTs in neurons, are collectively referred to as tauopathies.[172,173] The largest burden of tau pathology results from hyperphosphorylated tau leading to aggregation of tau filaments to form glial and neuronal tangles in dystrophic neurites.[174] Disturbances in the normal structural and regulatory functions in the cytoskeleton compromise axonal transport and thus contribute to synaptic dysfunction and neurodegeneration.[175,176] Aberrantly phosphorylated forms

of tau aggregate into insoluble fibrils that accumulate in the soma and dendrites of neurons where they have been associated with perturbed kinesin protein function and axonal transport and cytotoxicity.[177,178] In the brain, the assembly of these insoluble fibrils occurs rapidly and once they form they persist indefinitely.[179] In humans with AD, the percentage of neurofibrillary tangles increase with the progression of symptoms.[180] The electrophysiological properties of the neurons of mice are unaffected by the presence of neurofibrillary tangles.[181–183] Global neural network dysfunction occurs in mice with modest numbers of neurofibrillary tangles, and both memory deficits and neuron death can be dissociated from these fibrillar inclusions.[181–183] The lack and functionality of neurofibrillary tangles in these mouse models, led researchers in Dr. Karen Ashe's laboratory at the University of Minnesota to the discovery that neural network disruption can be caused by soluble forms of tau disrupting synaptic function.[184] Using a comparison of memory-impaired and unimpaired rTg4510 mice, a 35-kDa tau cleavage product was upregulated in impaired mice relative to unimpaired mice.[183–185] Completing a tryptic digest of the excised 35-kDa product found a peptide that corresponds to residues 299-314 suggesting that the fragmented tau product is generated by proteolysis at Asp314 and therefore was considered Δtau314.[184] Using a MEROPS protease database, it was found that only Caspases are able to cleave tau at Asp314.[184,186] Screening of all of the Caspases with tau, found that Caspase-2 is the only Caspase able to generate Δtau314 which leads to synaptic dysfunction and impaired cognition and is elevated 3-fold in AD brain.[184] rTg4510 mice dosed with anti-Caspase-2 morpholino oligonucleotides for 28 days exhibited 67% lower Δtau314 levels than mice dosed with vehicle-control, highlighting that Caspase-2 activity can be inhibited through use of targeted molecules.[184] Targeting Caspase-2 with a selective inhibitor should preserve cognitive function in tauopathies.

1.5.4 Caspase Structure and Comparison of Cas-2 and Cas-3

Structural biology efforts on several Caspases has led to valuable insight into the basis of caspase specificity and catalytic mechanism.[187] Crystal structures of many of the Caspases have

been solved: Caspase-1[188], Caspase-2[189], Caspase-3[190], Caspase-4 (PDBid: 6NRY, manuscript to be published), Caspase-6[191], Caspase-7[192], Caspase-8[193] and Caspase-9[194]. Active Caspases are comprised of a homodimer with the monomeric unit containing a large and small subunit.[160] Six anti-parallel β-strands of each monomer from a continuous 12-stranded β-sheet in the active enzyme that are surrounded by several α-helices and short β-strands which gives rise to the globular fold.[160] Each monomer has an active site which is formed by four protruding loops termed L1-L4 located at two opposite ends of the β-sheet (Figure 1.6).[160] L1 originates from the large subunit whereas the L3 and L4 loops come from the small subunit.[160] The L2 loop, which contains the catalytic cysteine, is cleaved between the large and small subunits into two segments.[160] L1 and L3 are relatively conserved in length and composition amongst the Caspases while L2 and L4 exhibit greater variability.[160] Selectively targeting Caspase-2 will require an in-depth understanding of L4 and the interactions peptide inhibitors make with L4.

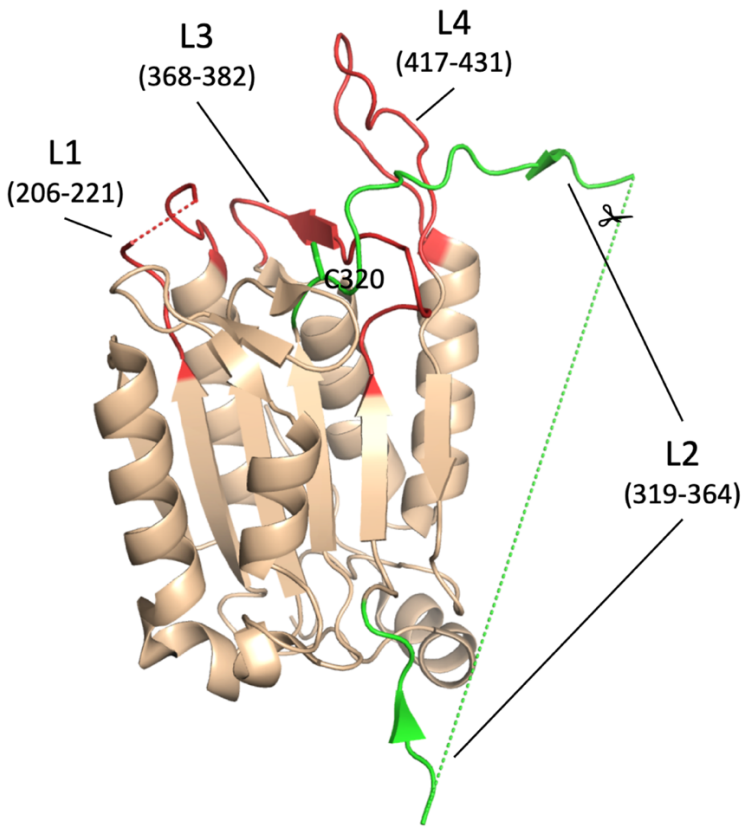


Figure 1.6 The Caspase Fold

Caspase-2 and Caspase-3 exhibit similar quaternary structure but vary in their substrate binding pocket which could be used for the basis of designing a selective molecule as described by Maillard et. al.[195] Important differences in the shape and size of the S2 subsite of Cas-2 and Cas-3 can be observed (Figure 1.7). Tyr204 in Cas-3 is replaced with Ala228 in Cas-2 and the Phe279 in Cas-2 moves approximately 3 Å towards the substrate binding site relative to the Cas-3 Phe256. The Cas-3 S2 subsite is lipophilic and forms a round, bowl-like shape that can preferentially bind shorter hydrophobic resides.[196] The S2 subsite of Cas-2 on the other hand forms a long groove that runs perpendicular to the active site cleft. Targeting this groove with either longer R-groups or bulkier amino acids (that would likely sterically clash with Tyr204 in Cas-3) could be an effective strategy in selectively targeting Cas-2. Another key difference between the two Caspases is the presence of an acidic amino acid, Glu52, that can engage P3 R-groups. The final difference is that

P5 substituents within the S5 subsite are tolerated and beneficial. Caspase-3 can be used in homology modeling of Cas-2 bearing in mind the structural differences described above for the successful generation of a potent and selective Cas-2 inhibitor. Once procedures and methodologies for Cas-2 yield co-crystal structures, Cas-2 and Cas-3, co-crystallized with the same molecules, can be directly compared for differences in binding modality and interactions with the active site.

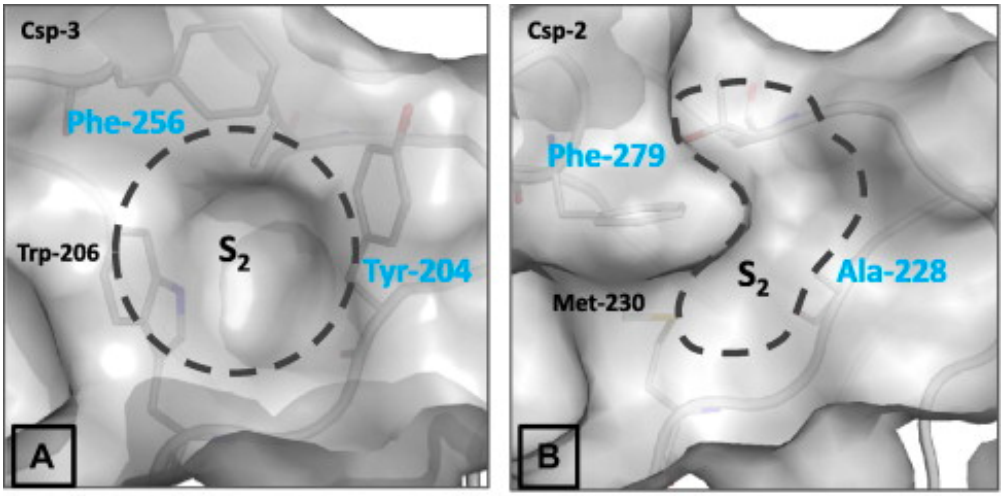


Figure 1.7 Comparison of the S2 Subsite in Cas-2 and Cas-3

Figure 1.7 was used with permission from: Maillard, M. C.; Brookfield, F. A.; Courtney, S. M.; Eustache, F. M.; Gemkow, M. J.; Handel, R. K.; Johnson, P. D.; Kerry, M. A.; Krieger, F.; Meniconi, M.; Sanjuan-Munoz, I.; Palfrey, J. J.; Park, H.; Schaertl, S.; Taylor, M. G.; Weddell, D; Dominguez, C. Exploiting differences in caspase-2 and -3 S2 subsites for selectivity: Structure-based design, solid-phase synthesis and in vitro activity of novel substrate-based caspase-2 inhibitors. *Bioorganic and Medicinal Chemistry* **2011**, *19* (1) 5833-5851. License number: 5123220524850

2. Structural characterization of the human cytosolic malate dehydrogenase I

The majority of this paper is largely reproduced (with the exception of the section involving the fragment screen and design of the *h*MDH1-pMCSG7 plasmid construct) with permission from: McCue, W.; Finzel, B. C. Structural characterization of the human cytosolic malate dehydrogenase I. *ACS Omega*. Manuscript Submitted.

Author Contributions: Dr. Shannon Kordus, a graduate student, at the time of the experiments, in Dr. Anthony Baughn's laboratory at the University of Minnesota, completed the microbiology surrounding the design of the *h*MDH1-pMCSG7.

2.1 Abstract

The first crystal structure of the human cytosolic malate dehydrogenase I (MDH1) is described. Structure determination at high resolution (1.65 Å) followed production, isolation and purification of human MDH1 using a bacterial expression system. The structure is a binary complex of MDH1 with only a bound malonate molecule in the substrate binding site. Comparisons of this structure with malate dehydrogenase enzymes from other species confirm that the human enzyme adopts similar secondary, tertiary, and quaternary structure, and that the enzyme retains much the same conformation even when nicotinamide adenine dinucleotide (NAD) is not bound. A comparison to the highly homologous porcine (*sus scrofa*) MDH1 ternary structures leads to the conclusion that only small conformational differences are needed to accommodate binding by NADH or other NAD mimetics, but that even NAD binding elements are somewhat flexible. Comparison of *h*MDH1 to the human mitochondrial malate dehydrogenase (*h*MDH2) reveals some key differences in the α7-α8 loop, which lies directly beneath the substrate binding pocket. These differences might be exploited in the structure-assisted design of selective small molecule inhibitors of hMDH1, an emerging target for the development of anti-cancer therapeutics.

2.2 Introduction

Malate dehydrogenases belong to the family of nucleotide-binding proteins referred to as NAD-dependent dehydrogenases or oxidoreductases.[105] This enzyme family includes the lactate dehydrogenases (LDHs), the liver alcohol dehydrogenases (LADHs), and the glyceraldehyde-3-phosphate dehydrogenases (GADPHs), among others.[105] Malate dehydrogenase reversibly converts malate to oxaloacetate with the use of NAD^+/NADH as a cofactor in the tricarboxylic acid cycle.[118] Eukaryotic cells contain two main isoforms which differ in their cellular compartmentalization and role in cellular processes: malate dehydrogenase II (MDH2) is found in the mitochondrial matrix where it is involved in the citric acid cycle, while malate dehydrogenase I (MDH1) is localized in the cytosol where it is important to the malate/aspartate shuttle of the Urea Cycle.[118]

Malate dehydrogenase I is overexpressed in a variety of cancers, and MDH1 amplification in human tumors is a common aberration that correlates with poor prognosis.[197] A hallmark of cancer cells is increased glucose consumption required for production of macromolecules necessary for growth and division.[120] Cytosolic NAD levels are independent of mitochondrial NAD levels involved in the electron transport chain.[122] Increased cytosolic concentrations of NAD are necessary to maintain the enhanced glycolysis of proliferating cancer cells, which has largely been attributed to the production of lactate through LDH activity.[120,122] Recently, it has been shown through the use of glucose isotopomer tracing in N5 cells that MDH1 supports LDH in the replenishing of cytosolic NAD.[125] The same study also showed that Jurkat cells with MDH1 knocked out (MDH1 KO.1 and MDH1 KO.2) show slower proliferation and glucose consumption than cells with functional MDH1.[125] This observation leads to the possibility that an MDH1 selective inhibitor, used either alone or in combination with LDH inhibitors, might slow tumor growth and cancer progression in patients.

Structural studies employing crystallography have been used to study malate dehydrogenase enzymes. Structures are known for mitochondrial (MDH2) enzymes from *E. coli*

(5KKA [130]; 3HHP [131]; 2CMD [132]), plants (1SMK [133]), and mammals (*sus scrofa*, 1MLD [134]). These MDH2 enzymes all share high sequence homology to human MDH2 (55-95%), but are distinct from the cytosolic (MDH1) enzymes that share lower homology (25-30%). Human MDH1 and MDH2 share only 26% sequence identity. Cytosolic MDH1 enzymes structures have been investigated from several species including bacteria (1BMD [135]; 4TVO [136]), plants (5NUE [137]). These enzymes share good sequence homology with the human MDH1 (50-62% identity). Collectively, this work has confirmed that the structures are quite homologous across all species. All MDH enzymes share a common Rossmann fold motif characteristic of other NAD(P) binding dehydrogenases [139], highly similar NAD+/NADH binding sites and mechanisms of catalysis.

Banaszak's group at the University of Minnesota completed some of the first high resolution crystallographic studies with porcine (*sus scrofa*) MDH1 that shares very high (95%) sequence identity with the human enzyme over 25 years ago [138,198], but to date no human MDH1 structure has been reported. Given renewed interest in *h*MDH1 as a possible therapeutic target, we have sought to obtain a crystal structure of the human enzyme in order to enable direct structure-aided design of an *h*MDH1 inhibitor. Here we report the first structure of the cytosolic human malate dehydrogenase I and compare it to both previously determined structures of other homologous cytosolic MDHs, and also to the structures of the human mitochondrial MDH (*h*MDH2). The monoclinic crystal form with malonate but no NAD^+/NADH bound provides a unique view of this emerging target for pharmaceutical development.

2.3 Materials and methods

2.3.1 Reagents

Full-length *h*MDH1 was a gift of Dr. Ameeta Kelekar (University of Minnesota, Department of Immunology, Minneapolis, MN). pGS-21a containing full-length *mdh1* was purchased from GenScript. All enzymes used for DNA digestion were purchased from NEB (Ipswich, MA). All DNA purification kits were purchased from Qiagen (Venlo, Netherlands).

Components necessary for protein production other than IPTG were ordered from VWR (Randor, PA). HisTrap HP column for protein purification was purchased from formerly GE Life Sciences, now Cytiva (Marlborough, MA). All other materials necessary for protein purification and isolation, crystallization, and determination of enzymatic activity were purchased from Fischer Scientific (Waltham, MA).

2.3.2 *h*MDH1-pMCSG7 construct and expression

Mini-prepped pcDNA 3.1-*mdh1* was digested with *Xho*1 and *Bam*HI at 37C for 30min, the enzymes were heat inactivated at 80C for 3min. pMCSG7 was also digested using *Xho*1 and *Bam*HI at 37C for 30min, the enzymes were heat inactivated at 80C for 3min. pMCSG7 was also treated with antartic phosphates at 37C for 30 min followed by a heat inactivation at 65C for 20min. The *mdh1* gene (roughly 1000bp) and pMCSG7 (roughly 5200bp) were separated using an agarose gel, cut out, and purified using a gel extraction kit. The products were ligated using T4 ligase in a 3:1 vector to insert ratio. The reaction was transformed into chemically competent DH5α cells and plated on LB+100µg/mL penicillin. The colonies were sequenced using the T7 and T7 term primers. Sequence verified plasmid was transformed into Rosetta2-pLysS (BL21 DE3) cells and plated on agar plates with Amp100 and Cm30. Expression parameters were varied to generate soluble protein. The inclusion bodies were solubilized in Urea Buffer (25 mM HEPES pH 7.2, 150 mM NaCl, 8 M urea, and 5% glycerol) and loaded onto a 5 mL HisTrap HP column equilibrated with Buffer A (25 mM HEPES pH 7.2, 150 mM NaCl, 10 mM imidazole and 5% glycerol). An SDS-PAGE gel analysis found a band corresponding to His-tagged *h*MDH1 (~36 kDa) in the flow-through fractions from the nickel column. Inability to separate *h*MDH1 during purification necessitated an exploration of solubility-tagged protein constructs to make purification easier.

2.3.3 Cloning, expression and purification of soluble His6-GST-*h*MDH1 fusion protein

The amino acid sequence corresponding to the full-length *h*MDH1$^{1-334}$ was codon optimized and cloned into a pGS-21a plasmid with additional TEV cleavage sequence using *bgIII* and *XhoI* restriction sites. The purchased plasmid was transformed into competent Rosetta2-pLysS (BL21 DE3) cells and plated on agar plates with Amp100 and Cm30. Colonies grew overnight and a single colony was selected and allowed to shake at 270 rpm overnight in LB media at 37°C. 1L LB media was inoculated with 3 mL of overnight culture and allowed to grow until OD_{600} reached 0.5-0.7. Culture was cooled down for one hour at 4°C prior to being induced with 1 mM IPTG. Cultures were placed in 20°C incubator and allowed to shake for 16 hours. Cells were harvested by centrifugation at 5000*g* for 15 minutes and stored at -20°C overnight. Cell pellet was solubilized in Buffer A (25 mM HEPES pH 7.2, 150 mM NaCl, 10 mM imidazole and 5% glycerol). To this solution was added lysozyme (final concentration: 1 mg/mL) and 1.5 µL Benzonase prior to sonication for 16 minutes (30 s on, 30 s off) at 30% attenuation. Lysed cells were distributed among 50 mL Beckman centrifuge tubes and centrifuged for 45 minutes at 45000*g*. Resulting supernatant was syringe filtered (0.45 µm) and loaded onto a 5 mL HisTrap HP column. Fusion protein was eluted from the column using a linear gradient of Buffer B (25 mM HEPES pH 7.2, 150 mM NaCl, 400 mM imidazole, and 5% glycerol). Like fractions were pooled together and TEV protease was added to 8% w/w before adding to dialysis cassette in TEV cleavage buffer (25 mM HEPES pH 7.2, 150 mM NaCl, 1 mM DTT, and 5% glycerol) at 4°C overnight. Contents of dialysis tubing was syringe filtered (0.45 µm) and loaded onto HisTrap HP column to separate *h*MDH1 from the His-tagged fusion protein. Flow through containing *h*MDH1 was pooled together and concentrated to ≤ 5 mL prior to being syringe filtered (0.22 µm) and loaded onto a Sephacryl S-100 column. Peaks from the size exclusion column were analyzed via SDS-PAGE and resulting *h*MDH1 was concentrated to 5 mg/mL in buffer supplemented with 10% glycerol. Aliquots were flash frozen before being stored in -80°C freezer for future use. Resulting yield from the preparation was

approximately 2.5 mg/L. Sequence was confirmed via crystallography and function confirmed via enzymatic activity assay.

2.3.4 Confirming enzymatic Activity

*h*MDH1enzymatic activity was assessed using a spectrophotometer to detect concentration of NADH at 340 nm. In short, to a vial was added NADH and oxaloacetic acid to either a vial with or without MDH1.[199] Loss of signal at 340 nm relative to the control sample (without enzyme) correlates to a decrease in the NADH/NAD^+ ratio as the enzyme is able to convert oxaloacetic acid to malate. Enzymatic activity was tested in only a qualitative manner to ensure structural integrity.

2.3.5 Crystallization

Protocols successfully used with *Sus scrofa* MDH1 served as a starting point for *h*MDH1 crystallization.[138] Microcrystals grew in 3-4 days by hanging drop vapor diffusion using precipitant consisting of 25-30% PEG 4000, 100 mM citrate buffer pH 6.5 and protein concentrated to 5 mg/mL in buffer containing 25 mM HEPES pH 7.4, 150 mM NaCl and 10% glycerol. Previous researchers identified malonate (pH range 4-8) as a good buffer for crystallization of LDHA, another member of the oxidoreductase family[200], so malonate was substituted for HEPES in crystal optimization. After several rounds of seeding and optimization, plate-like crystals (75 x 25 x 150 μm) reached full maturity within 48 hours at 20°C. Final conditions yielding monoclinic crystals were obtained by mixing equal amounts of protein with well solution containing 28% PEG 4000, 100 mM NaMalonate pH 7.2, and 0.15 mM ammonium acetate in 2 μL drops. Optimal cryo-conditions involved mother liquor plus 15% PEG 400.

2.3.6 X-ray data collection

Diffraction data was collected at IMCA-CAT beamline 17-ID at the Advanced Photon Source (APS), Argonne, Illinois, USA. Collection was completed at 100 K using radiation of wavelength 1.00 Angstroms and a Dectris Eiger2 9M detector. Data was processed using *autoProc* and re-scaled using *aP_scale* using R-factor (<.4), completeness (>90%), and I/sigma (>2) as

criteria, and a minimum of two of the three criteria were met in determining the proper resolution range.[201]

2.3.7 Structure solution and refinement

Due to the high sequence similarity between *Sus scrofa* MDH1 and *h*MDH1(96%), the *ss*MDH1 model, 5MDH (chain A) was used as the search model in molecular replacement with *Phaser*.[138,202] The monoclinic crystal was found to contain two protein chains in the asymmetric unit. Iterative rounds of refinement and model building were carried out using *Phenix*[203] and *Coot*[204]. Malonate was modeled into the corresponding electron density using standard geometry as in the CCP4 dictionary.[205] Refined structures were validated with *MolProbity* [206]. Table 2.1 summarizes statistics from the data collection and *Phenix* refinement for assessment of the quality of the structures.[207] Atomic coordinates and reflection data for *h*MDH1 with malonate bound in the substrate binding pocket have been deposited into the Protein Data Bank.[129] (Accession code: 7RM9) A second non-isomorphous crystal form was solved using the iterative process described above, but no electron density corresponding to malonate was observed. (Accession code: 7RRL) The majority of this chapter will focus on PDBid: 7RM9 because of its superior structure quality and apparent relevance. Section 2.4.6 compares the two non-isomorphous crystal forms.

2.3.8 Fragment screen using differential scanning fluorimetry (DSF)

A fragment screen of MDH1 was completed using differential scanning fluorimetry (DSF) on a BioRad CFX96 Well System (Hercules, CA) and previously established protocol.[208] To determine the optimal protein concentration, protein in varying concentration (0.3-0.01 mg/mL) was added to wells of a 96-well plate containing fixed concentration of Sypro-Orange (1:500 dilution of 5000x) and SEC buffer (25 mM HEPES pH 7.2 and 150 mM NaCl) to a final volume of 40 μL. The parameters of the scan were 0.5°C steps from 25-80°C with a 30-second hold at each temperature for proper measurement of the fluorescence signal. The resulting fluorescence signals were curve fitted using a Boltzmann model for non-linear regression in the curve fitting software,

Bio-Rad CFX 2.0, present on the instrument.[209,210] The optimal protein concentration was determined to be 0.2 mg/mL and this concentration was used for the rest of the subsequent assays. The cofactor, NADH, was tested in a 11-point dose-response (0.5-50 mM) DSF assay by adding NADH freshly dissolved in SEC buffer to the wells and using the same DSF parameters described above. Based off the results with NADH, 5 mM NADH was added to the well conditions for the DSF fragment screen. The Maybridge Ro3 Diversity Library of 1000 fragments in 96-well plate format were thawed at room temperature. 1 μL of 200 mM fragment was added to the corresponding well of the 96-well assay plate. The first and final column, were DMSO controls (2.5% v/v). The wells contained protein (0.2 mg/mL), Sypro-Orange (1:500 dilution of 5000x), NADH (5 mM), and fragment (5 mM). Same parameters were used and resulting data was fitted using the same Boltzmann model.

2.4 Results and discussion

2.4.1 Expression and purification of soluble protein

An initial construct design included an N-terminal His-tagged protein with a TEV cleavage site for straightforward purification and cleavage of the tag to isolate *h*MDH1 for crystallographic and enzymatic activity studies. Over-expression of *h*MDH1 was qualitatively achieved with small volume cultures at varying temperature (16-37 °C), IPTG concentrations (0.1-1 mM) and induction periods (1-20 hours). Following each trial induction, 1 mL culture aliquots were collected and centrifuged (13000*g*) and SDS-PAGE gel was used to compare the amount of ~36 kDa species relative to an uninduced sample. Highest over-expression was observed following induction with 1 mM IPTG overnight at 20 ºC.

The initial construct was designed for ease of purification by nickel affinity chromatography, but no soluble elution was observed with a trial purification from a 1 L culture. A band corresponding to the molecular weight of *h*MDH1 was observed in the insoluble fraction. In an attempt to solubilize the protein, expression was done at a lower temperature (16 ºC) with use

of a gradient of urea during lysis (1-8 M) to solubilize the protein from the insoluble fraction. Only at the highest concentration of urea was the protein successfully solubilized and therefore lysis buffer was supplemented with 8 M urea. Again, no protein with the appropriate molecular weight was found in any nickel flow-through fraction. Inability to solubilize and separate the protein from miscellaneous proteins after cell lysis compelled us to explore alternate constructs.

A modified pGS-21a plasmid was engineered with a TEV cleavage site and a codon optimized *mdh1* for bacterial expression was cloned into the modified plasmid. The rationale behind the engineered plasmid was to use codon optimization and a solubility-tag (GST) to increase likelihood of soluble protein after lysis. TEV cleavage site was included for easy cleavage of the solubility-tag from the isolated protein. The combination of the solubility-tag and codon optimization yielded soluble protein after cell lysis that was found to bind to the nickel column. TEV protease was able to cleave the GST-tag from *h*MDH1 which could then be further purified through a second nickel column (GST and His-tagged TEV protease stuck to the column while cleaved *h*MDH1 did not). Finally, Size Exclusion Chromatography was employed to ensure the highest possible protein purity for crystallographic and enzymatic studies.

2.4.2 Secondary, tertiary and quaternary structure of *h*MDH1

The *h*MDH1 crystal was found to have $P2_1$ symmetry with two protein molecules (the functional biological unit) in the asymmetric unit and the structure has been solved and refined to 1.65Å resolution (Table 2.1). Each monomer adopts the Rossmann fold characteristic of other NAD(P) binding dehydrogenases[139], comprised of nine α-helices and eleven β-strands conserved in all MDH structures (Figure 2.1). A lengthy insertion in the sequence of MDH1 enzymes between β8 and α7 gives rise to an additional beta pair ($\beta8^a$-$\beta8^b$) curled into the β8-α7 loop not present in MDH2 enzymes.

The site of NAD^+/NADH binding in this family of structures lies at the meeting of loops at the edge of the large parallel beta sheet within this fold. Loops β1-α1 and β2-α2 support the

adenosine diphosphate, β5-α5 cradles the nicotinamide nucleoside and β4-α4 provides a sort of cap that lays over all. Substrates (malate or oxaloacetate) bind just to the side of the nicotinamide base, pinched between alpha helices α7 and α8, held in precise position by H-bonds to universally conserved residues (Arg92, Arg98, Asn131 and His187).

Monomers assemble into biologically relevant homodimers, forming a helical bundle with α1 and α8 replicated across a non-crystallographic two-fold axis and winged with extensive contacts between α2 and α7 of opposite monomers (Figure 2.2). While there is no direct contact between active sites in the two monomers, previous work has shown that MDH1 functions as a dimer in solution and disruption of this interaction along the dimer interface could potentially perturb enzymatic function.[118] While the two monomers are in distinctly different crystallographic environments in our structure, they are quite similar overall (rmsd 0.192 Å). Crystal packing does influence conformational flexibility in this crystal form. The crystallographic B-factors of α-carbons throughout the structure are illustrated in Figure 2.3. Large B-factors are observed for residues 203-205 in subunit A. Amino acids in this region are modeled with lower occupancy (0.5) as positive difference peaks were observed in the $2F_o$-F_c map but little electron density was observed in the F_o-F_c map. The majority of the important β4-α4 loop (residues 92-99) is disordered in subunit B and cannot be modeled; there is no interpretable electron density for this loop. The rest of the structure is well ordered, with 98% of the amino acids show favorable torsional angles according to Ramachandran analysis with no outliers.

Table 2.1 Crystal Data and Refinement Statistics for Two Isomorphous *h*MDH1 Crystals

Data collection and processing		
PDB ID code	7RM9	7RRL
Diffraction source	APS 17-ID	APS 17-ID
Resolution range (Å)	38.82 - 1.65 (1.709 - 1.65)	46.01 - 2.05 (2.123 - 2.05)
Space group	P 1 21 1	P 1 21 1
a, b, c (Å)	51.503 84.312 77.652	67.875 62.593 74.322
α, β, γ (°)	90 91.223 90	90 91.219 90
Total reflections	144101 (15345)	76478 (7698)
Unique reflections	73789 (7932)	39102 (3914)
Multiplicity	2.0 (1.9)	2.0 (2.0)
Completeness (%)	92.56 (99.97)	99.45 (99.95)
Mean I/sigma(I)	13.90 (3.53)	8.53 (2.24)
Wilson B-factor (Å2)	17.86	30.62
Rmerge	0.02781 (0.1677)	0.04543 (0.3143)
Rmeas	0.03933 (0.2372)	0.06424 (0.4445)
CC1/2	0.999 (0.924)	0.997 (0.753)
CC*	1 (0.98)	0.999 (0.927)
Structure solution and refinement		
Reflections used in refinement	73786 (7931)	39092 (3914)
Reflections used for R-free	3685 (392)	1979 (208)
Rwork	0.1670 (0.1846)	0.1868 (0.2508)
Rfree	0.1927 (0.2213)	0.2374 (0.2993)
CC(work)	0.967 (0.932)	0.965 (0.824)
CC(free)	0.960 (0.862)	0.946 (0.735)
Number of non-hydrogen atoms	5396	5058
macromolecules	4994	4878
ligands	14	0
solvent	388	180
Protein residues	658	655
R.m.s. deviations		
Bonds (Å2)	0.007	0.007
Angles (°)	0.94	0.85
Ramachandran		
Favored (%)	98.62	97.53
Allowed (%)	1.38	2.47
Outliers (%)	0	0
Rotamer outliers (%)	0.92	0.97
Clashscore	2.67	3.38
Average B-factor (Å2)	19.89	29.77
macromolecules	19.48	29.75
ligands	20.76	N/A
solvent	25.1	30.26

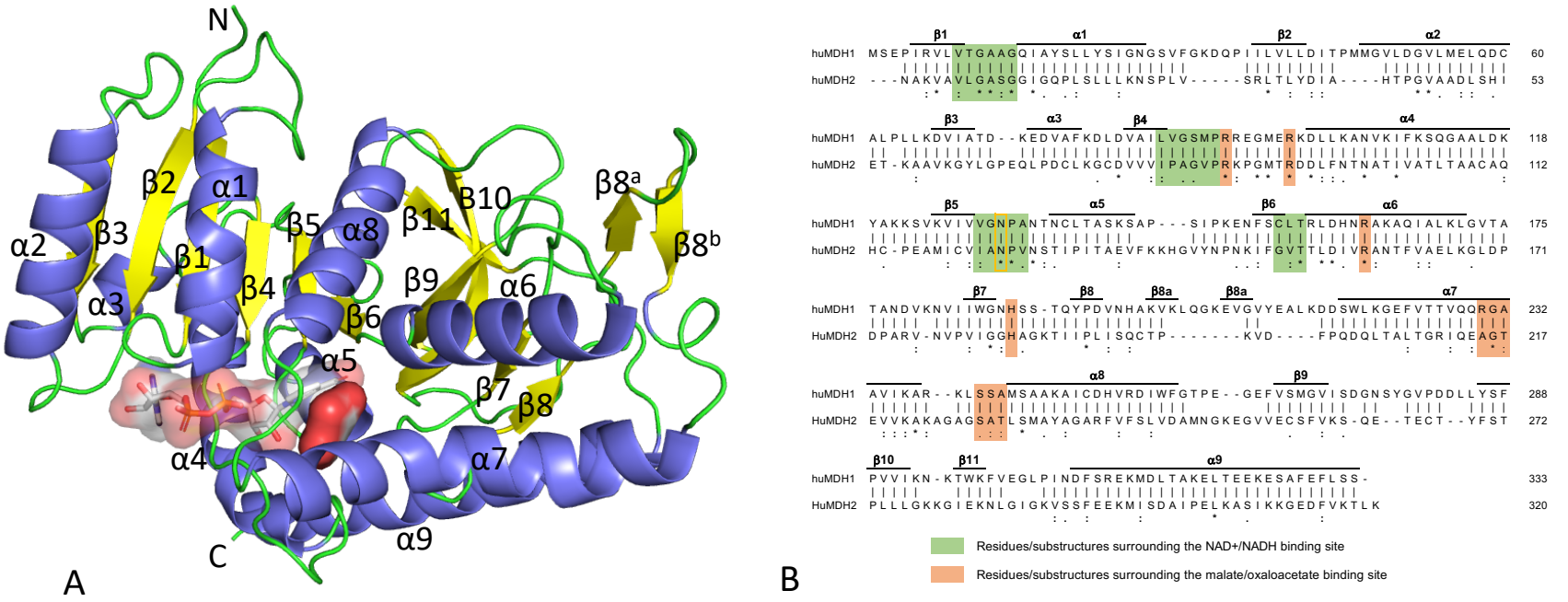


Figure 2.1 The Malate Dehydrogenase Fold and Sequence Comparison of *h*MDH1 and *h*MDH2

A) The Malate dehydrogenase fold and secondary structure assignments. B) A structure-based alignment of human MDH1 and MDH2. Vertical lines joining amino acids in the two sequences denote a 1:1 correspondence in position of residues in the structures of the two enzymes. Boxes identify cofactor and ligand binding motifs. Sequence conservation marks displayed (* :, .) are from a Clustal-Omega multiple sequence alignment of MDHs with known structures in the PDB.

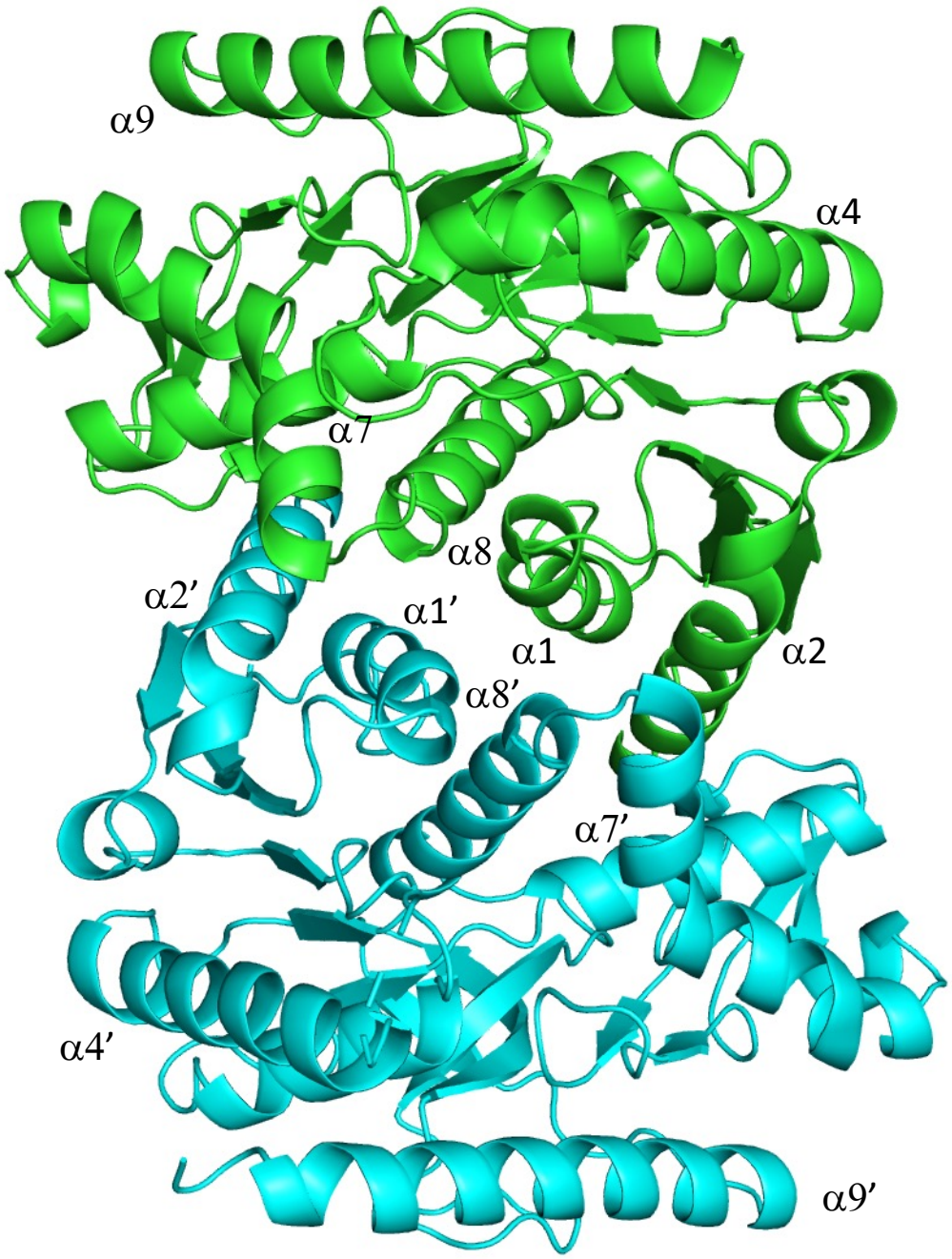


Figure 2.2 The Quaternary Structure of *h*MDH1 Viewed Along the 2-fold Axis

Chain A shown in green and chain B in cyan. Alpha helices are labeled to give greater clarity to the quaternary and tertiary fold of the protein with helices from chain A labeled as α# and chain B as α#'.

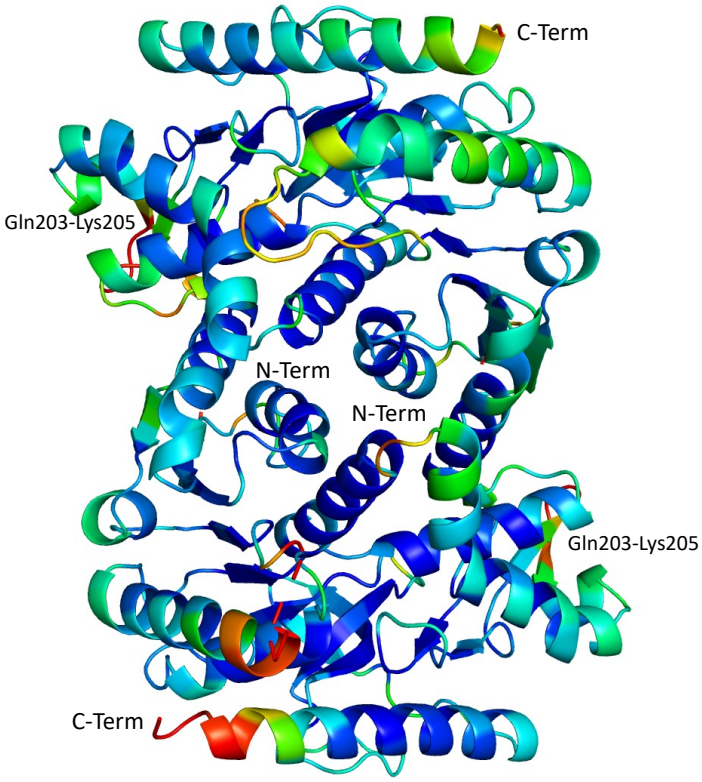


Figure 2.3 B-value Variability in *h*MDH1 Dimer

The protein backbone is colored by B-factors of CA atoms with a rainbow spectrum, minimum=12 (blue) and maximum=35 (red).

2.4.3 The NAD^+/NADH binding site

The structure reported here has neither NAD^+/NADH or a substrate bound. We have tried to form binary complexes with NADH, and ternary complexes with NADH and oxaloacetate by either soaking or co-crystallization under similar conditions, but these efforts have not yet been successful. The ternary complex of *Sus scrofa* (porcine) MDH1 (PDBid 5MDH) provides a good basis for modeling where cofactor and substrate should be when bound together in *h*MDH1, and for identifying conformational differences in the *h*MDH1 crystal form that might prevent co-factor binding. In the *ss*MDH1 complex (Figure 2.4A), hydrogen bonds from the backbone of the β1-α1 loop anchor the adenosine ribonucleotide, sidechains of Ser89 and Asn131 engage the nicotinamide

ribose, and His181 is H-bonded to the nicotinamide amide. An overlay of key NAD^+ binding segments (those highlighted in) from subunit A of *h*MDH1 and *ss*MDH1 gives a rms difference in backbone atoms of only 0.24 Å, confirming that the *h*MDH1 structure is poised to accept NAD^+/NADH with minimal conformational adaptation (Figure 2.4B). All interactions should be preserved accept the H-bond to Ser88, which is flipped to a different rotamer orientation in *h*MDH1 when NAD^+ is absent.

The binding elements in Subunit B are not well positioned for cofactor binding, however (Figure 2.4C). As mentioned above, residues 92-98 are completely disordered. While this loop makes little direct contact with NAD^+, it forms the floor of substrate binding pocket beneath the nicotinamide ring, and substrates typically H-bond to the cofactor. The Gly14-Gln15 amide bond in Subunit B is rotated roughly 90 degrees, so that the carbonyl is directed away from the phosphate of NAD to which it should H-bond. Asp42-Met45, and Gly88-Pro91 have all shifted into the NAD binding space, effectively shrinking the binding pocket.

We have examined crystal packing in the vicinity of NADH-binding sites in both subunits, but there is no simple explanation for these conformational differences; no intermolecular interactions exist to prevent subunit B from adopting a conformation similar to that seen in subunit A or *vice versa*. It is interesting that structural flexibility exists, as it affirms the possibility that small molecules might be identified that bind and stabilize conformations of *h*MDH1 that cannot support cofactor binding.

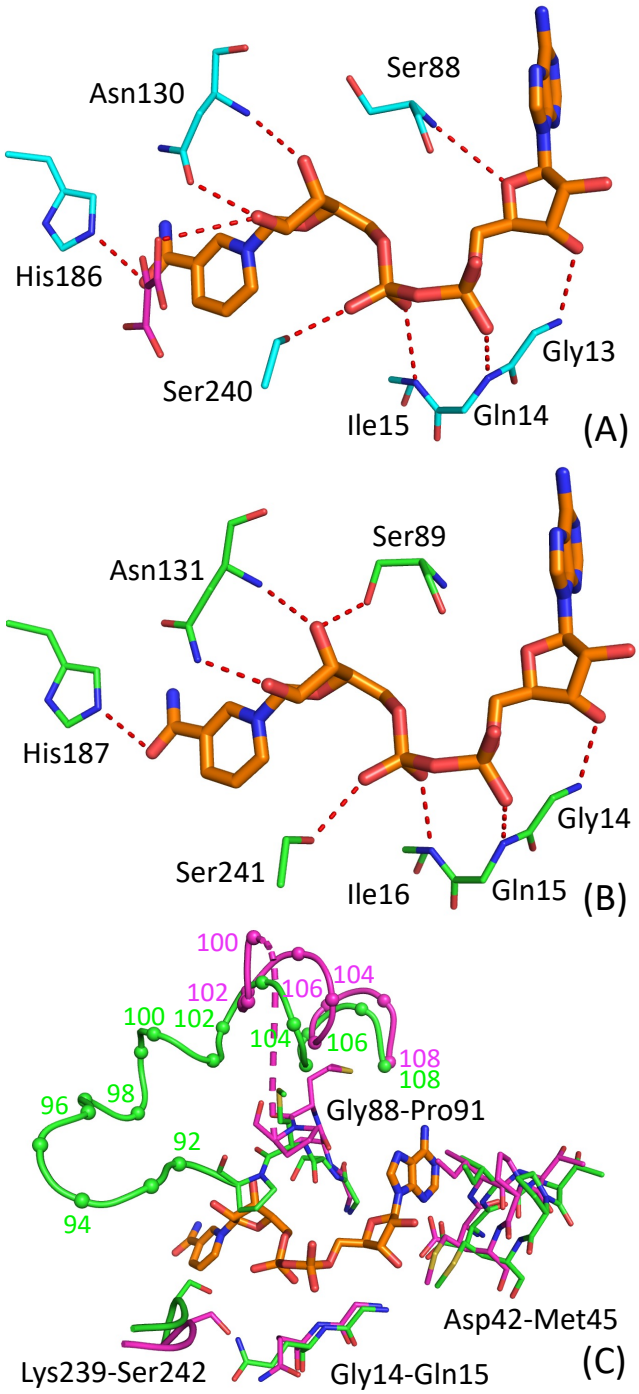


Figure 2.4 In-depth Look at tNAD Binding Interactions

(A) *ss*MDH1 hydrogen bonding network with tNAD. (B) Homology model of hMDH1 hydrogen bonding with tNAD. (C) Comparison of the cofactor binding pocket of hMDH1 subunit A (green) and hMDH1 subunit B (magenta).

2.4.4 Malonate lies in substrate binding pocket

Electron density present in the substrate binding pocket of subunit A can be attributed to malonate (Figure 2.5) a component of the buffer for crystallization. No comparable density is found in subunit B, which, as described above, is less well-ordered. Arg92 and Arg 98, which figure prominently in positioning substrate molecules (see below), are completely disordered in subunit B.

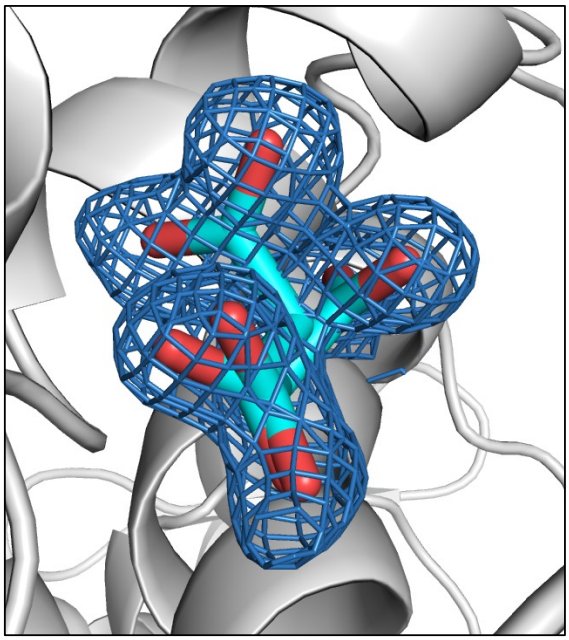


Figure 2.5 Omit Map of Malonate Bound in *h*MDH1 Structure.

Map mesh vectors (blue) lying more than 1.8 Å from the ligand are hidden.

Malonate has been modeled in the *h*MDH1 complex in two different conformations with half occupancy (Conformer A and Conformer B). The hydrogen bonding stabilizing each conformation is shown in Figure 2.6A. In both conformations, one carboxylate in this symmetric molecule is positioned opposite the plane of the Arg92 guanidinium to which it is H-bonded. In Conformation A, the other carboxylate interacts with Arg162, Arg98, and-Ser242. In Conformation B, this carboxylate is rotated to hydrogen bond with His187 and the secondary amine in Arg98. These interactions collectively mimic those seen by more natural substrate analogs that are one

carbon longer and can make both sets of interactions, such as those observed in the porcine MDH1 complex with α-ketomalonate (PDBid 5MDH) (Figure 2.6A). Interestingly, malonate binds in this pocket even in the absence of the NAD^+ co-factor. In the α-ketomalonate complex, one oxygen of the carboxylate is able to hydrogen bond with an exocyclic oxygen of the nicotinamide ribose. We could find no other example of a malate dehydrogenase structure in the PDB that includes a bound substrate analog in the absence of NAD^+/NADH.

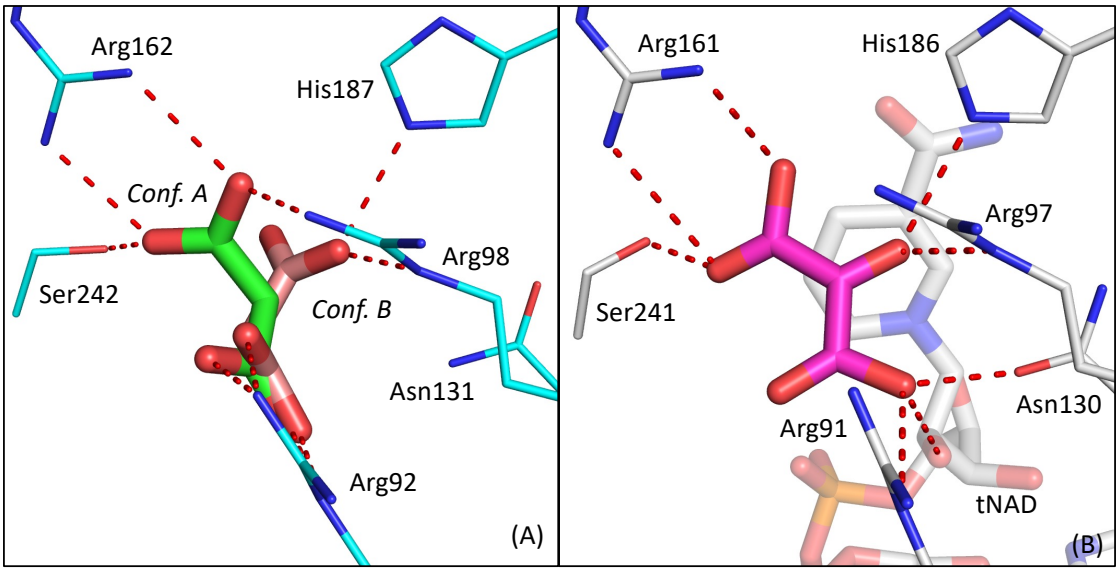


Figure 2.6 Comparison of Hydrogen Bonding Network Between Malonate in *h*MDH1 and α-ketomalonate in *ss*MDH1

(A) Hydrogen bonding to malonate by *h*MDH1. (B) Hydrogen bonding to α-ketomalonate by *ss*MDH1 (PDBId: 5MDH)

2.4.5 Differences between human cytosolic and mitochondrial MDH

Any successful targeting *h*MDH1 for therapeutic purposes will require that some degree of selectivity be achieved over the mitochondrial MDH (*h*MDH2). While it has never been described in detail, a high resolution (1.9 Å) structure of a ternary complex of *h*MDH2 was solved in 2006 by the Structural Biology Consortium and deposited in the PDB with accession code 2DFD. Its existence affords an opportunity for detailed comparison to the *h*MDH1 enzyme. The structure-

based alignment of residues with PDBeFold[211] results in the pairwise alignment of sequences illustrated in Figure 2.1B and an RMSD of paired backbone atoms of 2.0 Å. Despite the higher RMSD, *h*MDH2 preserves the fold and all the cofactor binding motifs with high homology.

One feature stands out as potentially relevant to prospects for selective inhibitor design in the comparison of *h*MDH1 and *h*MDH2: the insertion of two extra residues into the α7-α8 loop. This loop passes directly under the substrate binding pocket (D-malate in the *h*MDH2 ternary complex) and contributes the surface that underpins the bound substrate (Figure 2.7). In addition to the insertion of two residues that results in a local shift in registry, there are specific sequence differences ($^{h\text{MDH1}}\text{Ile}_{235}$ to Val; $^{h\text{MDH1}}\text{Ser}_{242}$ to Ala; $^{h\text{MDH1}}\text{A}_{243}$ to Thr) that alter the shape of the substrate binding pocket. There are also likely significant differences in the flexibility and dynamics of these two loop variants in response to ligand binding. While the malate binding pocket is small, these differences may afford an opportunity for selective small molecule inhibitor design.

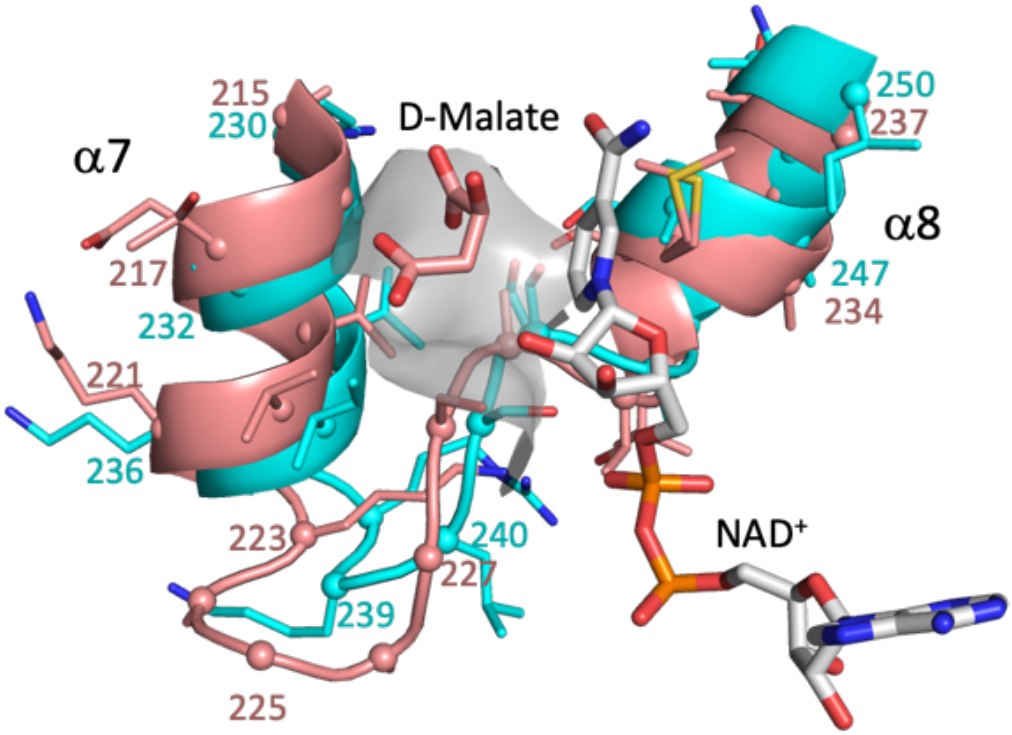


Figure 2.7 Comparison of the α7-α8 loop in *h*MDH1 vs. *h*MDH2

*h*MDH2 (salmon), NAD^+ and D-malate are from the ternary *h*MDH2 complex (PDBid: 2DFD). *h*MDH1 is shown in cyan. Overlay results upon superposition of NAD binding substructures identified in Figure 2.1B (RMSD 0.79Å). The portion of this loop in contact with bound malate is illustrated with the fragment of surface.

2.4.6 Comparison of non-isomorphous *h*MDH1 crystal forms

Two non-isomorphous crystals forms were solved that share the same space group ($P2_1$) with a dimer in the asymmetric unit, but vary in the unit cell parameters, specifically the length of the *a*- and *b*-axes that alters intermolecular packing significantly (Table 2.1). Superimposing chains A of both structures shows a global RMSD difference of 0. 217 Å, but closer examination exposes large conformational differences between the active site loop (Figure 2.8). Gly95 moves approximately 10 Ångstroms between the ligand bound and empty active site conformations. This more open conformation is similar to one observed in *E. Coli* MDH2 when NAD^+/NADH was not

present (3HHP [131]). Further investigation is needed as to why one crystal yielded the closed (NAD^+/NAPH)-bound conformation when another crystal grown under the same conditions did not.

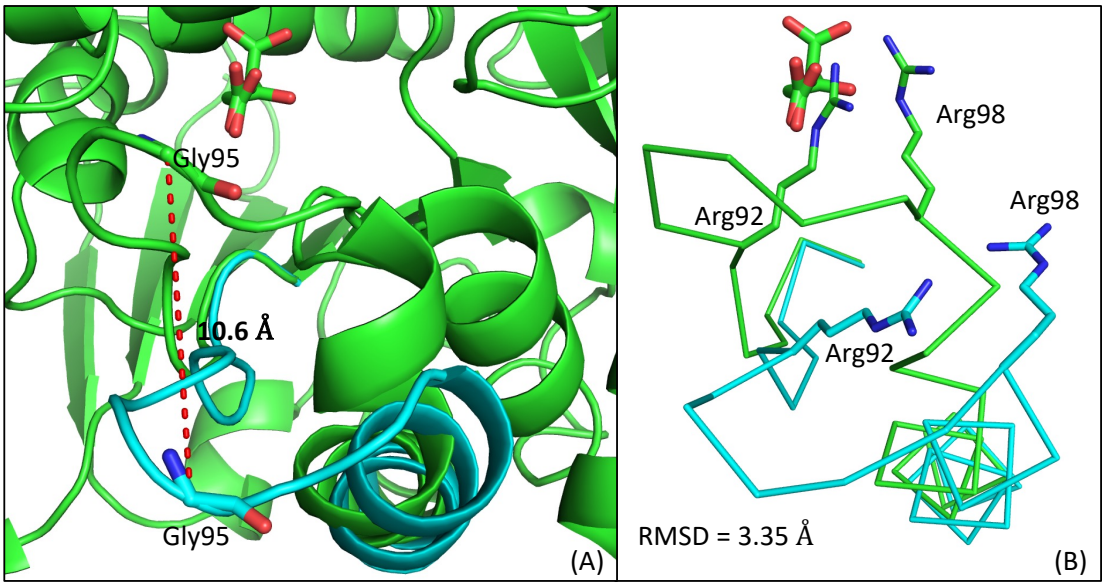


Figure 2.8 Examination of the difference in the α4-β4 loop that cradles NAD+/NADH in two crystal forms of *h*MDH1.

Shown are cartoon (A) and Cα ribbons (B) of chain A in Form I (green) (PDBid: 7RM9) vs chain A in Form II (cyan) (PDBid: 7RRL).

2.4.7 MDH1 unfolding exhibits biphasic transition

A fragment screen was conducted on *h*MDH1 to begin the drug discovery process using differential scanning fluorimetry (DSF) as it is fast, robust, and inexpensive. Prior to beginning the actual screening campaign, conditions for the DSF assay needed to be optimized. Protein concentration was varied from 0.3-0.01 mg/mL with an optimal concentration of 0.2 mg/mL being used as it was the lowest concentration of protein that gave a robust fluorescence signal (>30,000 RFU). An issue with the melting curves, was the presence of two denaturation transitions (~50 ºC and 58 ºC) made evident in the first derivative plot (Figure 2.9). Biphasic curves can originate from a variety of different variables: multiple protein species in solution, poorly or improperly folded

protein, and denaturation of the quaternary structure, to name a few.[212] Various conditions were tested to decrease the biphasic nature of MDH1 including a buffer and pH screen and the addition of additives, such as glycerol, to attempt to stabilize the protein fold. None of the conditions tested were able to decrease the biphasic nature of the curves and therefore the conditions were left unchanged and changes in either of the transitions would be analyzed in the fragment screen. The effect of NADH on the melting curve was tested in a dose-response (1-50 mM final concentration in the well) assay and NADH was found to cause a 6ºC positive shift in the second observed transition of the unfolding temperature of MDH1 relative to apo protein (64 ºC vs. 58 ºC), (Figure 2.10). The maximal change in denaturation temperature was observed at 2.5 mM and therefore the fragment screen would be completed at 5 mM NADH to ensure proper saturation. Presence of NADH did not decrease the biphasic nature of hMDH1. Since malonate was found to bind to *h*MDH1 and mimic several of the same hydrogen bonds observed of α-ketomalonate in *ss*MDH1, it was tested in the DSF to determine the effect on protein denaturation. Malonate was found to not effect either of the transitions of the biphasic denaturation of *h*MDH1 (Figure 2.11).

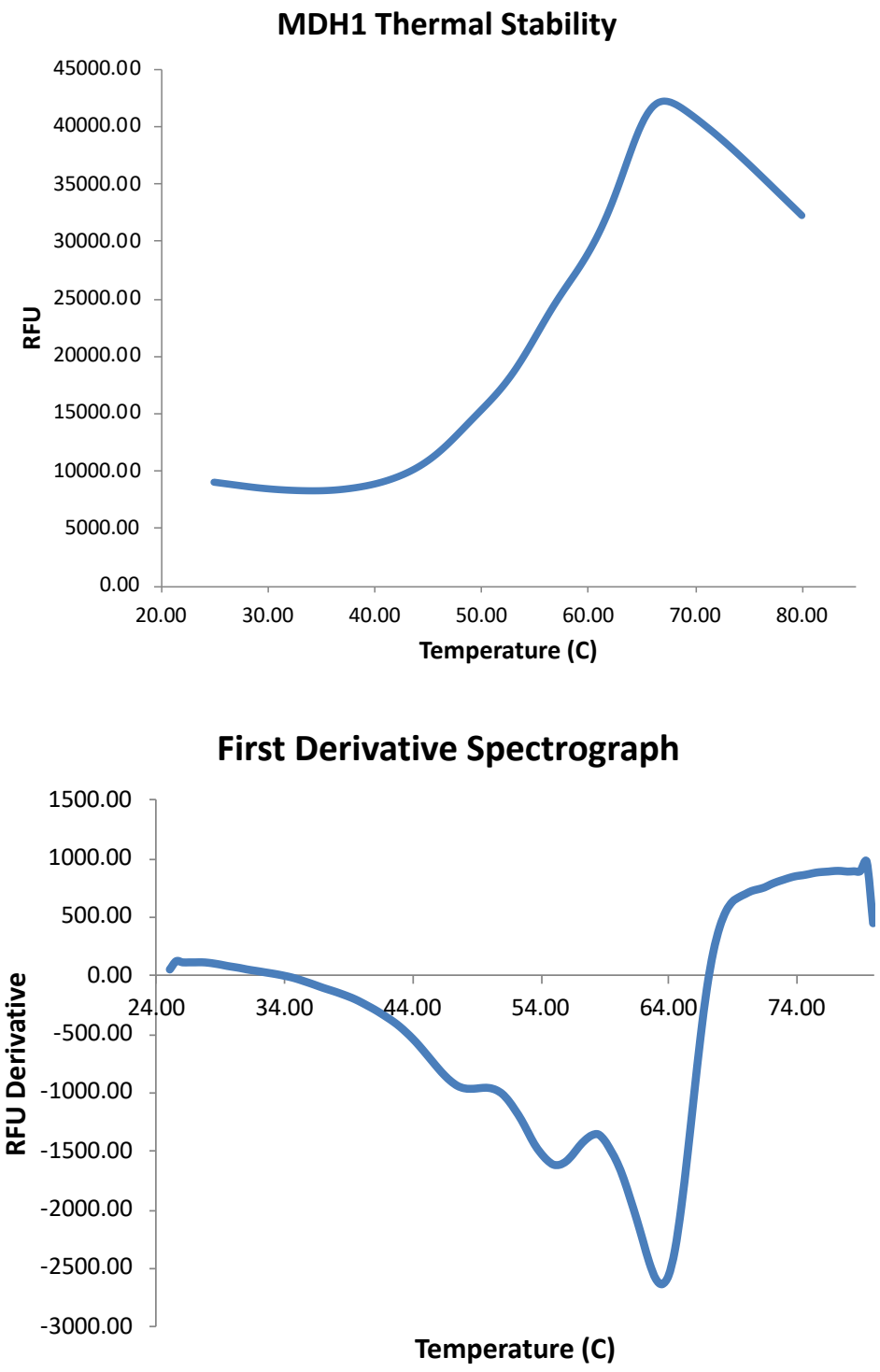


Figure 2.9 Biphasic Transition of MDH1 Denaturation

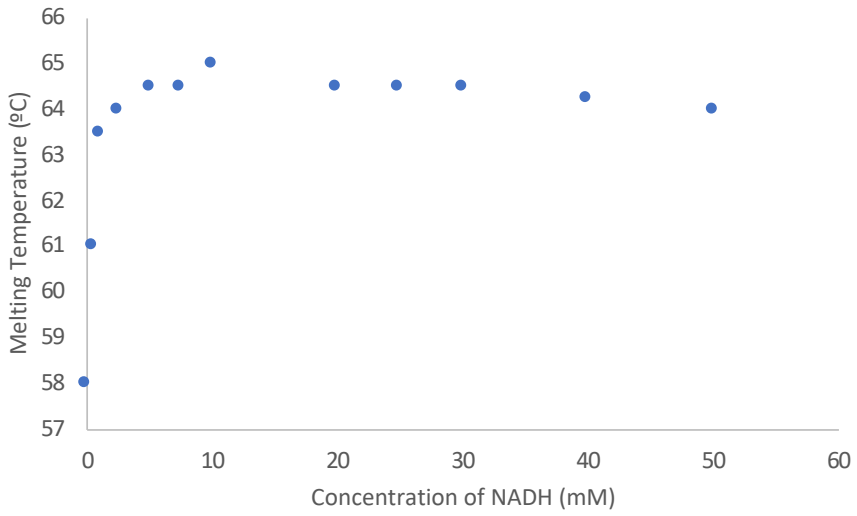


Figure 2.10 NADH Dose-Response in DSF Assay

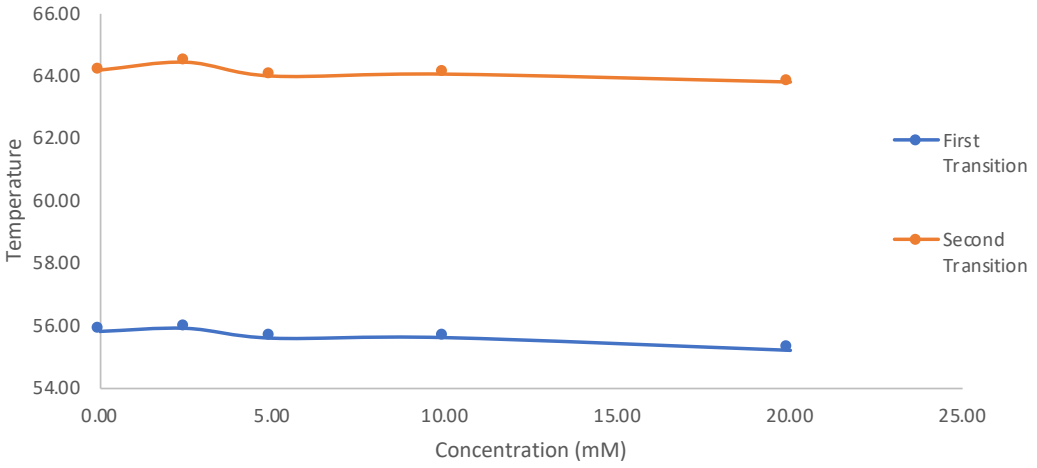


Figure 2.11 Malonate Does Not Affect MDH1 Protein Stability

2.4.8 Fragment screen produced no hits to follow-up

A fragment screen was conducted on a Maybridge Ro3 Diversity Library of 1000 fragments. The library is Rule of 3 compliant and therefore all fragments abide by the following physico-chemical properties to allow for successful lead discovery: molecular weight is <300 Da, number of hydrogen bond acceptors is ≤3, number of hydrogen bond donors is ≤3, and ClogP is ≤3.[213] Screening of the compounds was completed in a 96-well assay format with each fragment at 5mM/well.

Since the denaturation of *h*MDH1 is biphasic, compounds which caused an increase or decrease (>1ºC) in either of the transitions relative to DMSO were considered hit fragments in the initial screen. In drug discovery, traditionally, positive shifting molecules and fragments are prioritized as they increase the protein stability, but no chemical matter was discarded during initial screening so both positive and negative shifting fragments were considered hits.[214] 18 fragments were found to cause at least ±1ºC shift. To accurately confirm the fragments, initial vials from Maybridge were used. Of the 18 fragments, 3 were unavailable for re-assay, 5 did not give a transition (fluorescence signal did not increase with temperature), 7 were unable to replicate the original thermal shift, and 3 caused a negative shift to give an initial hit rate of 0.3% which is a log-fold worse initial hit rate than expected for a DSF fragment screen.[215] None of the original hits reproduced the T_m shift of the original screen and therefore there are no hits to follow-up. The fragments came from vials that were plated onto source plates using the Echo in the Institute for Therapeutic Drug Discovery here at the University of Minnesota. The source plates are thawed and refrozen each time a fragment screen is conducted and therefore could be susceptible to degradation. The low hit rate is indicative that the use of DSF to discover fragment molecules as inhibitors of *h*MDH1 is not the correct approach and other approaches such as SPR should be pursued in future work.

2.5 Conclusions and outlook

The first over-expression, isolation, and purification of human MDH1 using a bacterial expression system has been described in detail. Further, the first *h*MDH1 crystal structure has been determined at 1.65 Å resolution with only a small molecule (malonate) bound in the active site. While the structure confirms that the enzyme is very similar to the previously reported porcine structure with which it shares 95% sequence identity, this new structure without bound NAD co-factor provides some novel insights into the conformational flexibility of the enzyme. The β4-α4 loop that cradles NAD^+/NADH in active enzymes adopts a conformation similar to that needed to bind NAD^+ in one crystallographic environment but is largely disordered in another. While

crystallization can exaggerate the importance of conformations stabilized by crystal packing, it is likely that this range of motion is accessible to the protein in solution as well, A more open conformation suggestive of extensive protein flexibility has previously been observed in an *E. coli* MDH2 apo structure, (3HHP[131]), but not in any MDH1 structure. This flexibility may leave the enzyme susceptible to inhibition by small molecule dinucleotide mimetics that trap the enzyme in an inactive conformation.

The comparison of the *h*MDH1 structure to *h*MDH2 is also revealing. Distinct differences in these two enzymes in the vicinity of the substrate binding pocket might be exploited in the discovery of small molecules that inhibit human MDH1 with selectivity over the mitochondrial MDH2. This selectivity will almost certainly be a desirable attribute of any agent put forward for clinical evaluation of the therapeutic potential of MDH1 inhibitors in the treatment of cancer. The monoclinic *h*MDH1 crystal form reported here may be a particularly useful tool for use in future crystallographic fragment screening, specifically because of its extensively open and empty active site.

3. Examination of a Series of Pentapeptide Inhibitors of Caspase-3 Using Crystallography

Author Contributions:

The pentapeptides were synthesized by Merlin Bresinsky and Dr. Steffen Pockes in the Institute of Pharmacy at the University of Regensburg. The synthesis of the molecules is not detailed, but the structures are included. The molecules were used in the enzymatic activity assay and for co-crystallization with Caspase-3. Jessica Strasser in the Institute for Therapeutics Discovery and Development (ITDD) at the University of Minnesota completed the Caspase-3 enzymatic activity assay and associated data analysis. The methods section detailing the assay (3.2.3) was written by Jessica Strasser.

A background on tauopathies, the Caspase family of proteins, the role of Caspase-2 in neurodegenerative diseases, the general Caspase structure and a comparison of Caspase-2 and Caspase-3 is included in Chapter 1.

3.1 Introduction

The grant which supported this work (5R01AG062199-03) has the broad, long-term objective of the development of specific caspase-2 inhibitors as neuro-therapeutic agents.[216] Caspase-3 (Cas-3) and Caspase-2 (Cas-2) are both apoptotic caspases that differ in their activation but share very similar protein folding (Figure 3.1) and therefore Cas-3 can serve as a surrogate model of Cas-2. The goal of Cas-3 crystallography is to be able to use resulting co-crystal structures in homology modeling and comparison of peptide binding interactions with the protein to increase potency and selectivity in Cas-2 inhibitor design. Crystallographic efforts in our laboratory have led to structure solutions of various Caspase-3 co-crystal complexes while progress on Caspase-2 has been much slower due to inadequate protein supplies and inability to grow any crystalline material. The pentapeptides with Cas-3 are the only current structural depiction of the peptides (no small-molecule crystals or Cas-2 co-crystals), and therefore, visualization of the movement of the peptides and shifts in the loops of Cas-3 can guide future Cas-2 inhibitor design. This chapter of the thesis will solely focus on the Caspase-3 co-crystal structures and not discuss any potentiation

for inhibitor design for Caspase-2 based on homology modeling or future co-crystal structure comparison. Here we describe 10 novel pentapeptide-Cas-3 co-crystal structures and compare the binding modality and conformational differences observed in the L3 and L4 loops and examine structure activity relationships.

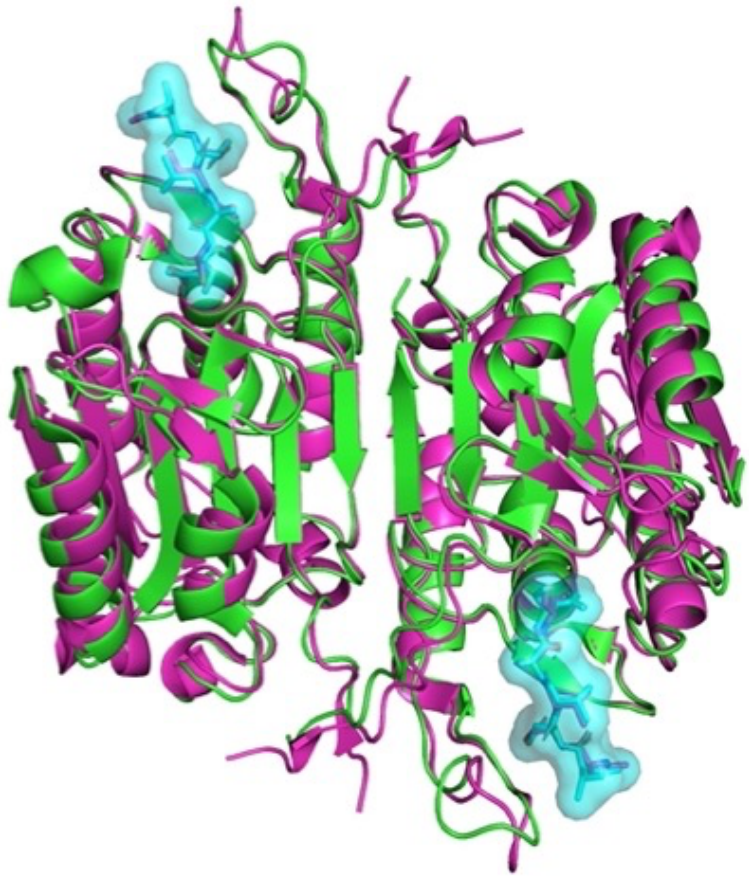


Figure 3.1 Overlay of Cas-2 (PDBid: 3R5J, Magenta) and Cas-3 (PDBid: 2H65, Green)

3.2 Methods

3.2.1 Reagents

The pET23b vector containing human Caspase-3^{29-277} was gifted by Dr. Michelle Arkin at UCSF. Bacto™ yeast extract and Bacto™ tryptone for protein expression were purchased from VWR (Radnor, PA). 5 mL HisTrap HP column for nickel column chromatography and 5 mL HiTrap Q HP column for ion exchange chromatography used for protein purification were purchased from formerly GE Healthcare, now Cytiva (Chicago, IL). Isopropyl β-d-1-

thiogalactopyranoside (IPTG) and components for protein purification including buffer components, syringes, and pipettes were purchased from Thermo Fisher Scientific (Waltham, MA). Crystal trays and cover slips for hanging drop vapor diffusion crystallography were ordered from Hampton Research (Aliso Viejo, CA). AFC fluorogenic substrate, Ac-DEVD-AFC, was purchased from Bachem (Torrance, CA). The 10 novel pentapeptides, supplied as 20 mM stock solutions in DMSO, were synthesized by Steffen Pockes at the University of Regensburg and has affiliation with the University of Minnesota. The synthesis will not be discussed in the context of this chapter.

3.2.2 Caspase-3 Expression, Isolation and Purification

The expression and subsequent purification were accomplished using a previously published protocol for caspases.[217] *Escherichia coli* Rosetta2 pLyss DE3 cells were transformed with the plasmid and plated onto an LB Amp100 Cm30 agar plate. Single colony was selected and added to 2 mL of 2X TY media supplemented with Amp100 and Cm30 and placed in an incubator at 37ºC for 8 hours as the primary culture. 0.5 mL was used to inoculate 50 mL of 2X TY media supplemented with the same antibiotics for overnight growth. The following morning, 20 mL of the overnight culture was added to a fresh 1 L of 2X TY media and allowed to grow until mid-log phase, OD_{600}=0.5-0.7, was achieved. Cultures were subsequently induced with 0.2 mM isopropyl-b-D-thiogalactosidase (IPTG) and incubated at 30ºC for 5 hours with 270 rpm shaking. Cell pellet was recovered through centrifugation at 4000*g* for 10 minutes to which 30 mL of resuspension buffer (100 mM Tris pH 8.0, 100 mM NaCl) prior to storage at -20ºC overnight. The cell pellet solution was removed from the freezer and thawed at 37ºC prior to sonication for 8 minutes (30 seconds on followed by 30 seconds off) in an ice bath. The lysed cells were spun down at 20000*g* for 30 minutes and the resulting supernatant was syringe filtered (0.45 μM) prior to loading onto a HisTrap nickel column pre-equilibrated with the resuspension buffer. Bound lysate was washed with a washing buffer (100 mM Tris pH 8.0, 500 mM NaCl) and eluted from the column using a linear gradient of elution buffer (100 mM Tris pH 8.0, 100 mM NaCl, 400 mM imidazole). SDS-

PAGE gel analysis was performed on fractions from the column, and like fractions corresponding to Cas-3 were pooled together and diluted 1:3 (protein:buffer) with a 100 mM Tris pH 8.0 buffer without salt. The diluted protein was loaded onto an ion-exchange column and eluted from the column with buffer containing 100 mM Tris pH 8.0 with ~175 mM NaCl. Presence and purity of Caspase-3 was analyzed using SDS-PAGE and an enzymatic activity assay. Like fractions were pooled together and diluted to 0.5 mg/mL with added 10% glycerol prior to flash freezing in liquid nitrogen and storage at -80°C. Typical yield for the expression, isolation, and purification was approximately 3 mg/L.

3.2.3 Caspase-3 Enzymatic Activity Assay

The enzymatic assay was performed in a 96-well plate format and Caspase-3 inhibition by compounds were measured in fluorometric assays. Caspase-3 was expressed, isolated, and purified as described above. Enzyme was diluted to 2.07 nM/well in buffer containing 100 mM HEPES pH 7.0, 150 mM NaCl, 0.1% CHAPS, 1.5% sucrose, and 10 mM DTT. Ninety-six and a half microliters of enzyme in buffer were added per well in a black Corning 3356 96-well assay plate. Test compounds were serially diluted in dimethyl sulfoxide (DMSO) and plated in duplicate or triplicate in a Corning 3357 transfer plate. Test compound was added to assay plates in 1 µL aliquots per well using a BiomekFX and mixed 10 times. Compound and enzyme mixture was incubated at 37°C for 5 minutes. The BiomekFX was then used to add and mix 2.5 µL of the AFC substrate, Ac-DEVE-AFC (0.4 mM), in DMSO from a transfer plate to the assay plate for a total assay volume of 100 µL in the assay plate. Fluorescence from free AFC was read at 37°C every 5 minutes over an hour using a Clariostar reader (λ_{ex} = 400 nm, λ_{em} = 505 nm). GraphPad Prism (v9) was used to calculate the IC_{50} by fitting the dose-response data with four parameter variable slope nonlinear regression.[218] The 40 minute time point was reported, consistent with reported literature.[195,219–221]

3.2.4 Co-crystallization of Cas-3 with Select Pentapeptide Inhibitors

Two different previously published well conditions were used to generate co-crystals of Caspase-3 and select inhibitors.[222,223] 500 μL of Caspase-3 (100 mM Tris pH 8.0, ~175 mM NaCl, and 10% glycerol) at 0.5 mg/mL was incubated with 500 μM inhibitor for 30 minutes on ice to allow for covalent bond formation. The solution was then concentrated to 4 mg/mL and protein concentration was tested using the A_{280} on the nanodrop. Cas-3-ligand crystals were generated by mixing equal amounts of protein with well solution containing either Condition 1: 15% PEG 6000, 5% glycerol (v:v), 100 mM sodium citrate pH 5.3, 10 mM DTT, and 30 mM NaN3 or Condition 2: 16% PEG 6000, 5% glycerol, 100 mM sodium citrate pH 6.5, and 10 mM DTT. Plate-like crystals grew within 24-48 hours and were cryoprotected with well solution supplemented with 10% PEG 6000 prior to flash freezing.

3.2.5 X-ray Data Collection

Diffraction data was collected at IMCA-CAT beamline 17-ID at the Advanced Photon Source (APS), Argonne, Illinois, USA. Collection was completed at 100 K using radiation of wavelength 1.00 Ångstroms and a Dectris Eiger2 9M detector. Data was processed using *autoProc* and re-scaled using *aP_scale* using I/sigma (>2), R-factor (<.4), and completeness (>90%) as criteria, with a minimum of two of the three criteria being met to determine the proper resolution range.[201]

3.2.6 Data Processing and Structural Refinement

The structure of Cas-3 co-crystallized with Ac-VDVAD-CHO, PDBid: 2H65, was used as a search model for molecular replacement with *Phaser*.[202,224] The monoclinic and orthorhombic crystal forms were found to contain a homodimer in the asymmetric unit with the monomer being composed of the small (p12) and large (p17) subunits. Iterative rounds of refinement and model building were carried out using *Phenix*[203] and *Coot*[204]. Pentapeptide inhibitors were modeled into the electron density using *Coot* and the covalent bond between Cys163 and the P1 aspartic acid

was constrained by adding a link record to the PDB file prior to subsequent round of refinement. The bond between Cys163 (atom name SG) and P1 (atom name C) was restrained to 1.5Å with a sigma of 0.5 and the thioester carbon center geometry was restrained to be planar. Refined structures were validated with *MolProbity*.[206] Table 3.1 was compiled with data from the data collection and use of *Phenix* for assessment of structure quality.[207] Atomic coordinates and reflection data for 10 pentapeptide co-crystal structures with Cas-3 were deposited into the Protein Data Bank and coordinates will be made available upon pulication.[129] PDB accession codes are included in Table 3.1.

Table 3.1 Data Processing and Refinement Statistics for Cas-3

Data collection and processing					
PDB ID	**7RN7**	**7RN8**	**7RN9**	**7RNA**	**7RNB**
Ligand	Ac-VD(Aly)VD-CHO	Ac-VD(Orn)VD-CHO	Ac-VDFVD-CHO	Ac-ITV(Dab)D-CHO	Ac-VDRVD-CHO
Resolution range	43.28 - 2.4 (2.486 - 2.4)	43.28 - 2.4 (2.486 - 2.4)	42.35 - 1.67 (1.73 - 1.67)	43.2 - 1.9 (1.968 - 1.9)	56.22 - 1.75 (1.813 - 1.75)
Space group	P 1 21 1	P 1 21 1	P 21 21 21	P 1 21 1	P 63
a, b, c (Å)	50.509 68.041 82.493	50.509 68.041 82.493	67.472 85.195 97.61	50.055 66.166 83.217	129.825 129.825 60.415
α, β, γ (°)	90 90.608 90	90 90.608 90	90 90 90	90 90.926 90	90 90 120
Total reflections	41458 (4026)	41458 (4026)	127065 (10909)	68436 (6860)	112174 (11468)
Unique reflections	21706 (2160)	21706 (2160)	64162 (5853)	41044 (4176)	56317 (5772)
Multiplicity	1.9 (1.9)	1.9 (1.9)	2.0 (1.9)	1.7 (1.6)	2.0 (2.0)
Completeness (%)	98.36 (98.18)	98.36 (98.18)	97.13 (89.45)	95.42 (97.80)	95.92 (98.68)
Mean I/sigma(I)	4.44 (2.19)	4.44 (2.19)	14.75 (2.46)	11.02 (2.30)	17.08 (4.43)
Wilson B-factor	30.39	30.39	21.75	30.54	22.28
R-merge	0.1175 (0.3766)	0.1175 (0.3766)	0.02208 (0.2335)	0.03228 (0.419)	0.0208 (0.1348)
R-meas	0.1661 (0.5326)	0.1661 (0.5326)	0.03122 (0.3302)	0.04565 (0.5926)	0.02942 (0.1906)
CC1/2	0.965 (0.718)	0.965 (0.718)	0.999 (0.921)	0.999 (0.731)	0.999 (0.943)
CC*	0.991 (0.914)	0.991 (0.914)	1 (0.979)	1 (0.919)	1 (0.985)
Structure solution and refinement					
Reflections used in refinement	21693 (2161)	21693 (2161)	64072 (5853)	41013 (4176)	56310 (5772)
Reflections used for R-free	1107 (133)	1107 (133)	3259 (283)	1991 (186)	2816 (284)
R_{work}	0.1878 (0.2272)	0.1878 (0.2272)	0.1831 (0.2943)	0.1945 (0.2817)	0.1668 (0.1945)
R_{free}	0.2515 (0.3158)	0.2515 (0.3158)	0.2099 (0.3383)	0.2310 (0.3268)	0.2018 (0.2507)
CC(work)	0.946 (0.877)	0.946 (0.877)	0.969 (0.907)	0.963 (0.867)	0.969 (0.916)
CC(free)	0.919 (0.655)	0.919 (0.655)	0.962 (0.844)	0.935 (0.789)	0.959 (0.844)
# of non-hydrogen atoms	3915	3915	4130	3945	4150
macromolecules	3804	3804	3835	3815	3884
ligands	6	6	6	0	6
solvent	105	105	289	130	260
Protein residues	473	473	472	475	478
R.m.s. deviations					
Bonds ($Å^2$)	0.008	0.008	0.008	0.008	0.007
Angles (°)	1.09	1.09	0.93	1	0.95
Ramachandran					
Favored (%)	96.48	96.48	98.04	98.26	97.85
Allowed (%)	3.52	3.52	1.96	1.74	2.15
Outliers (%)	0	0	0	0	0
Rotamer outliers (%)	0	0	0	0.24	0
Clashscore	5.18	5.18	4.99	5.94	3.37
Average B-factor	30.12	30.12	24.23	34.56	24.96

Table 3.1 Data Processing and Refinement Statistics for Cas-3, Continued

Data collection and processing					
PDB ID	**7RNC**	**7RND**	**7RNE**	**7RNF**	**7RNG**
Ligand	Ac-VDVVD-CHO	Ac-VDPVD-CHO	Ac-YKPVD-CHO	Ac-VDKVD-CHO	Ac-ITAKD-CHO
Resolution range	42.36 - 1.933 (2.002 - 1.933)	40.9 - 2.15 (2.227 - 2.15)	63.38 - 2.73 (2.828 - 2.73)	33.54 - 2.111 (2.187 - 2.111)	32.31 - 2.55 (2.641 - 2.55)
Space group	P 21 21 21	P 1 21 1	P 21 21 21	P 1 21 1	P 63
a, b, c (Å)	67.248 85.221 97.621	50.306 68.609 81.803	68.103 84.16 96.333	50.34 67.081 82.183	129.226 129.226 60.642
α, β, γ (°)	90 90 90	90 90.45 90	90 90 90	90 91.137 90	90 90 120
Total reflections	82243 (8347)	57475 (5421)	30366 (2948)	59043 (6105)	37935 (3760)
Unique reflections	41470 (4144)	29662 (2924)	15215 (1471)	30149 (3101)	19022 (1882)
Multiplicity	2.0 (2.0)	1.9 (1.9)	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)
Completeness (%)	96.76 (98.69)	97.35 (96.79)	99.43 (99.59)	95.41 (99.81)	99.88 (100.00)
Mean I/sigma(I)	7.14 (2.34)	12.15 (4.49)	10.85 (2.95)	9.26 (2.28)	9.28 (3.73)
Wilson B-factor	25.05	26.45	42.48	30.85	40.86
R-merge	0.07418 (0.2447)	0.03825 (0.1936)	0.03869 (0.2103)	0.04503 (0.3021)	0.06626 (0.2657)
R-meas	0.1049 (0.3461)	0.05409 (0.2738)	0.05471 (0.2974)	0.06368 (0.4273)	0.0937 (0.3758)
CC1/2	0.974 (0.942)	0.998 (0.931)	0.998 (0.946)	0.998 (0.861)	0.987 (0.776)
CC*	0.994 (0.985)	0.999 (0.982)	1 (0.986)	0.999 (0.962)	0.997 (0.935)
Structure solution and refinement					
Reflections used in refinement	41305 (4144)	29638 (2924)	15156 (1471)	30132 (3102)	19012 (1882)
Reflections used for R-free	2060 (187)	1421 (145)	737 (76)	1497 (179)	908 (97)
R_{work}	0.2343 (0.2855)	0.1687 (0.2033)	0.1904 (0.2870)	0.1782 (0.2066)	0.1774 (0.2431)
R_{free}	0.2840 (0.3121)	0.2218 (0.2778)	0.2535 (0.3852)	0.2368 (0.2630)	0.2416 (0.2911)
CC(work)	0.932 (0.927)	0.965 (0.935)	0.961 (0.890)	0.966 (0.924)	0.955 (0.834)
CC(free)	0.905 (0.855)	0.945 (0.806)	0.920 (0.796)	0.926 (0.814)	0.939 (0.833)
# of non-hydrogen atoms	4038	4021	3798	3977	3899
macromolecules	3822	3870	3765	3847	3811
ligands	6	0	0	6	0
solvent	210	151	33	124	88
Protein residues	476	475	471	472	474
R.m.s. deviations					
Bonds ($Å^2$)	0.008	0.007	0.008	0.008	0.008
Angles (°)	0.99	1.37	0.95	0.95	0.95
Ramachandran					
Favored (%)	98.71	97.41	96.95	97.61	98.05
Allowed (%)	1.29	2.16	3.05	2.39	1.95
Outliers (%)	0	0.43	0	0	0
Rotamer outliers (%)	0	0.24	0.97	0	0.48
Clashscore	5.4	3.92	7.91	4.57	5.41
Average B-factor	26.79	28.92	42.48	34.62	38.91

3.3 Results and Discussion

3.3.1 Confirming Protein Purification and Crystallographic Methods

Expression, isolation, purification, and co-crystallization with peptides of human Cas-3 using a bacterial expression system has been achieved by previous researchers.[225] The target was new to the Finzel laboratory and therefore the first objective was optimizing the protein expression and determining the optimal conditions for crystallography of the purified protein. Inoculation of the large culture from the overnight culture was slow to reach mid-log phase (OD_{600}=0.5-0.7) as it took approximately 8 hours (relative to 3 hours observed for *h*MDH1 discussed in Chapter 2) and yielded a small cell pellet following centrifugation (~2 g/L). Caspase is a proteolytic enzyme and therefore could be self-cleaving leading to cellular death resulting in slower growth. To potentially circumvent this adverse effect and increase the speed of cell growth, an additional growth step, suggested in a different protocol, was added prior to inoculating the large culture and the media was switched to 2X TY media that has been used for other Caspases.[217] The additional growth step decreased the time needed to reach mid-log phase and increased the size of the resulting cell pellet (~5 g/L). Isolation of Cas-3 involved using the His-tag and purification followed with an ion-exchange column to yield pure protein that was confirmed in the enzymatic activity assay by Jessica Strasser to be equal or slightly more active than the purchased protein used in the Walters' laboratory.

Initial efforts in crystallography involved screening previously published conditions for Caspase-3 co-crystals. The initial attempts involved incubating the published protein concentration (4 mg/mL or ~140 μM) with a 20-fold molar excess of ligand (2.8 mM) on ice for 30 minutes to ensure covalent linkage prior to setting up crystal trays with varying conditions. Previous published conditions with this incubation step and Ac-VDVAD-CHO did not lead to any crystal growth. Using a 50 mM stock solution of peptide dissolved in DMSO and targeting 2.8 mM final peptide concentration, meant that the ratio of protein:ligand needed to be ~ 17:1 resulting in a final DMSO

concentration (v:v) of 5.5%. Previous enzymatic activity work had shown that the Caspase family of enzymes lose enzymatic activity with increasing concentration of DMSO.[195,219–221] To decrease the amount of concentration of DMSO, ligand was diluted to 0.5 mg/mL and incubated with the same molar excess of peptide inhibitor and incubated on ice for 30 minutes. Following incubation, the protein-peptide complex was concentrated to 4 mg/mL using a centrifugal filter unit. The incubation at dilute concentration and subsequent concentration decreased the DMSO concentration and led to crystal growth. Two conditions, described in the methods as Condition 1 and Condition 2, were found to generate plate-like crystals that grew within 24-48 hours (Figure 3.2). The fast growth observed generated thin plates that were able to gain a little more three-dimensional character with time. Efforts to slow the growth included varying protein, PEG, and DTT concentration and varying the pH of the citrate buffer, but no combination greatly increased the 3-D character of the crystal or changed the morphology. Freshly solubilized DTT was found to be critical to crystal growth. To establish Cas-3 co-crystallization in the Finzel laboratory, co-crystal complexes with previously published ligands Ac-VDVAD-CHO (PDBid: 2h65), Ac-LDESD-CHO (PDBid: 3edq), and Ac-DEVD-CHO (PDBid: 1pau) were grown and diffraction data collected. The structures were found to contain electron density for the peptide inhibitor bound to Cys163 of each monomer of the asymmetric unit (Figure 3.3) and mapped well onto previously published structures with equal or improved resolution (Figure 3.4). The crystallographic data and refinement statistics for these structures are not included in the chapter as the structures are not novel and were meant only as a proof of concept to show collaborators that we were ready for new peptides to structurally characterize.

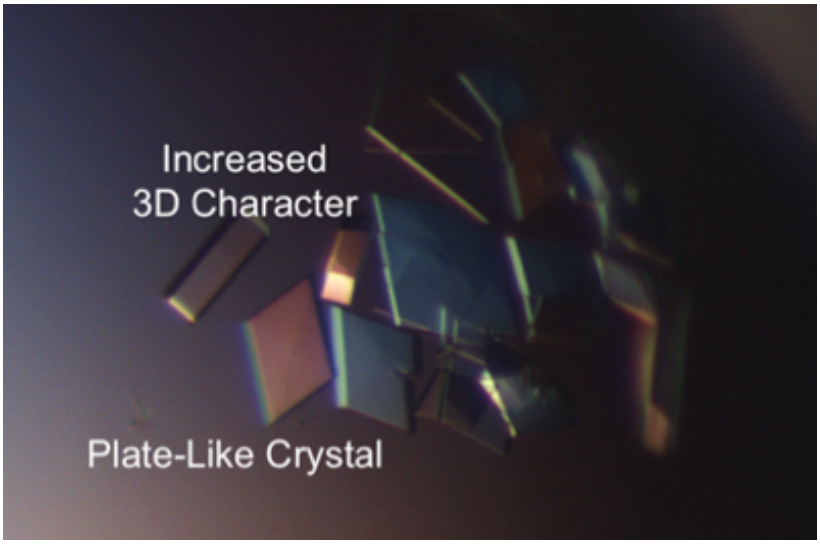


Figure 3.2 Microscope Image of Cas-3 Crystal Morphology

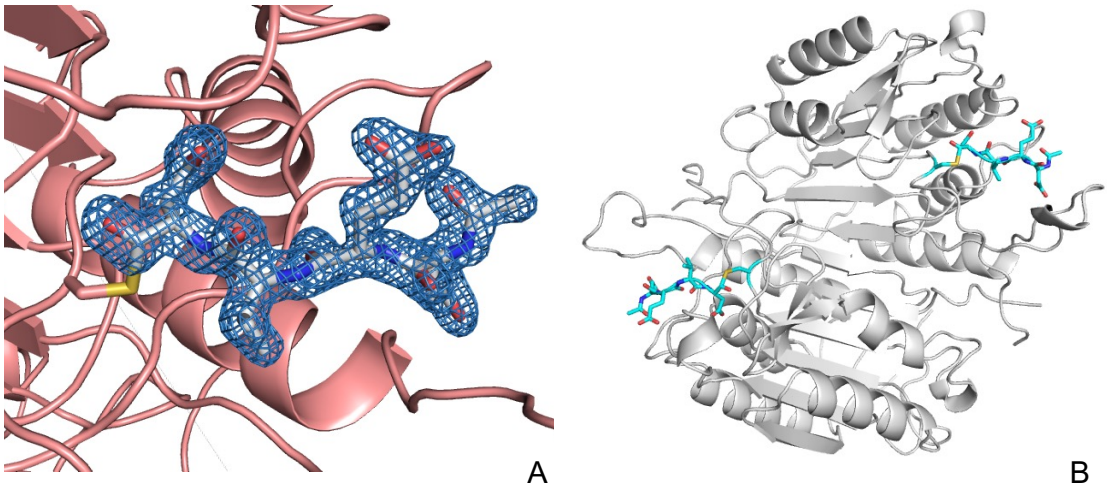


Figure 3.3 Structure of Ac-DEVD-CHO and Caspase-3

A) Omit map of covalently bound Ac-DEVD-CHO. Map mesh vectors lying more than 1.8 Å from the peptide are hidden. B) Dimeric structure of Caspase-3 with two bound instances of the ligand shown in cyan.

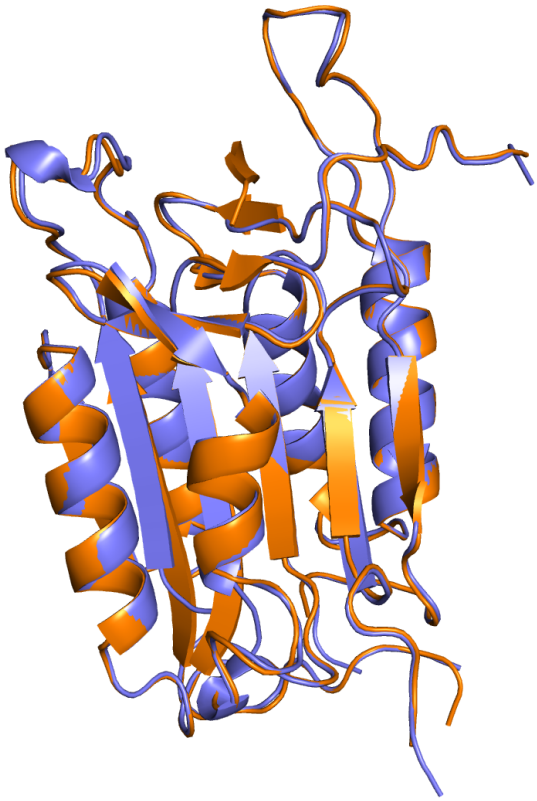


Figure 3.4 Overlay of Ac-DEVD-CHO Structure to Previously Published Structure

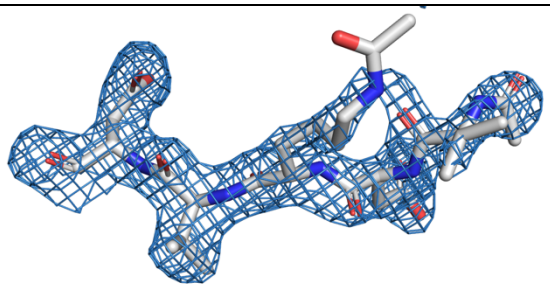
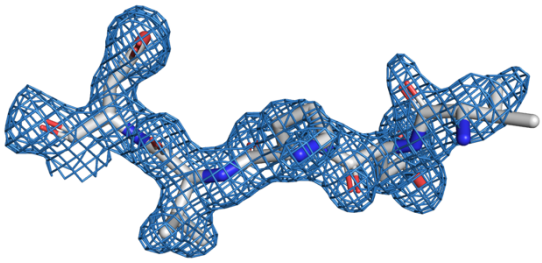
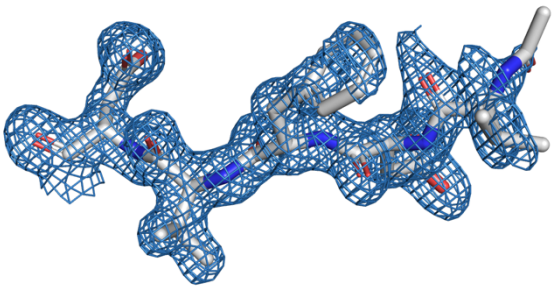
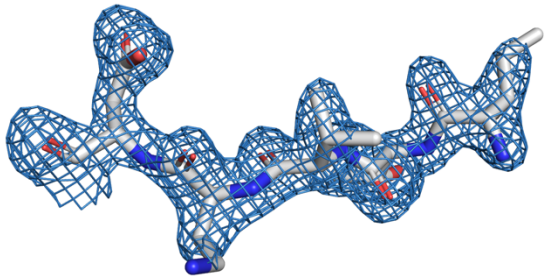
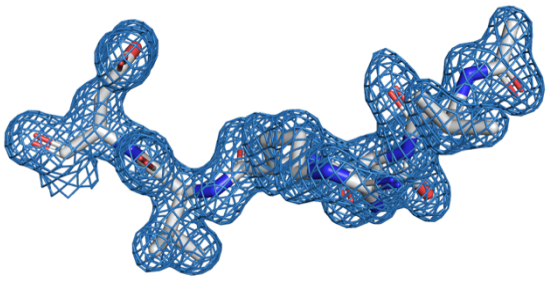
Overlay of the co-crystal structure of Caspase-3 with Ac-DEVD-CHO at a higher resolution shown in purple with the previously published structure with the same ligand (PDBid: 1pau) in orange. Global RMSD for superposition is 0.286 Å.

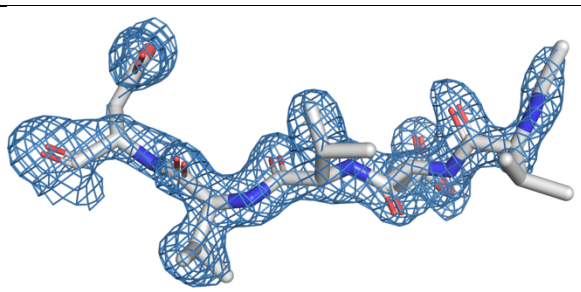
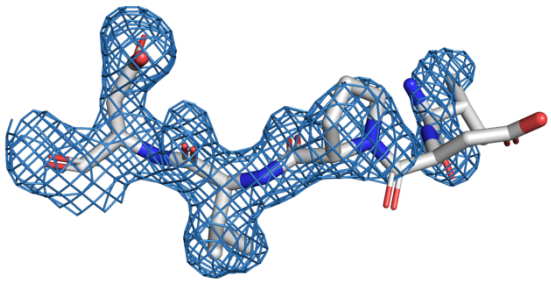
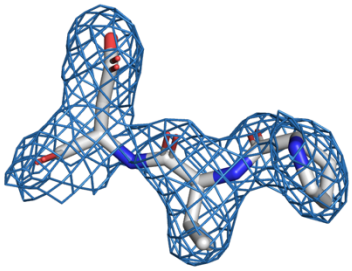
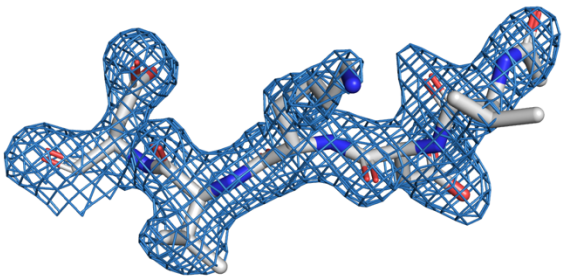
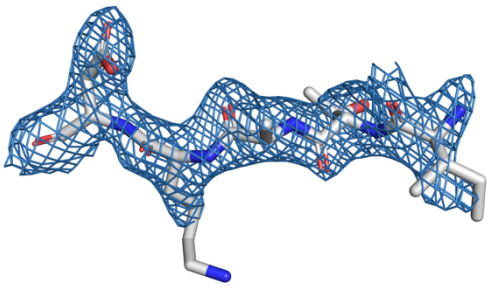
3.3.2 Crystallographic Data of 10 Novel Caspase-3 Co-crystal Structures

A total of 10 crystal complexes were generated, solved, and deposited for Caspase-3 co-crystallized with pentapeptides (Table 3.1). The structures are all monoclinic, orthorhombic, or hexagonal and exhibit one of three space groups ($P2_1$, $P2_12_12_1$, $P6_3$). Resolution of the structures varies from 1.66-2.73 Ångstroms, but the resolution of the structures does not have any correlation on the space group or K_i of the pentapeptide. The R_{work} for the majority of the structures is less than 0.2 which is better or comparable than structures published at the same resolution (Ac-VDVVD-CHO, PDBid: 79NC, has a higher than expected R_{work} of 0.238).[129] The majority of the R_{free} values

are within 5% of the R_{work} values showing that the models were properly fitted into the working set of reflections.[226] While the structures vary in their resolution, space group and R-factors, they have several observed characteristics that are maintained between them. The unit cell is composed of a dimer of the basic biology subunit, p17 and p12 subunits. The p17 subunit and p12 subunit combine to form the expected loops (L1, L2, L3, and L4) that form the substrate binding pocket. All structures contain two instances of the pentapeptide that has the aspartic acid covalently bonded to Cys163. The bond formation was forced in the model building and refinement in *Coot* and *Phenix*, respectively, but electron density supports the disulfide bond (Table 3.2). The pentapeptides are labeled Ac-XXXXX-CHO and the formation of the covalent bond between Cys163 and the aspartic acid at P1 causes the cleavage and loss of the CHO functional group. The molecule names were kept in this format for deposition into the PDB (even though the CHO is not present) to retain the name of the molecule that was synthesized and co-crystallized with Cas-3. The thioester carbon geometry was forced trigonal planar, as this is the expected geometry, using planarity restraints for refinement in *Phenix*. The N-terminus of the pentapeptides have the nitrogen acetylated, but electron density for many of the structures does not support the modification and therefore the acetylation group was not modeled for said structures.

Table 3.2 Ki and Omit Maps for Novel Pentapeptides

Ligand	Ki (nM)[1]	Omit Map[2]
Ac-VD(Ack)VD-CHO	26	
Ac-VD(Orn)VD-CHO	58	
Ac-VDFVD-CHO	24	
Ac-ITV(Dab)D-CHO	1441	
Ac-VDRVD-CHO	120	

Ac-VDVVD-CHO	34	
Ac-VDPVD-CHO	3519	
Ac-YKPVD-CHO	36,000	
Ac-VDKVD-CHO	77	
Ac-ITAKD-CHO	910	

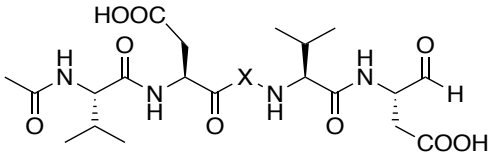
[1]Ki values are from Caspase-3 enzymatic activity assay. Work completed by Jessica Strasser.

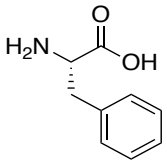
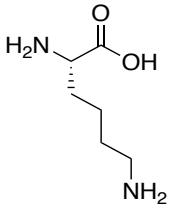
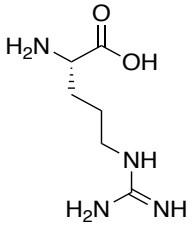
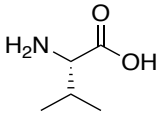
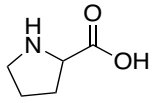
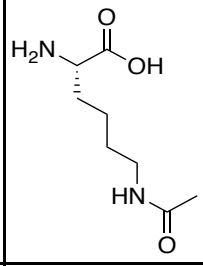
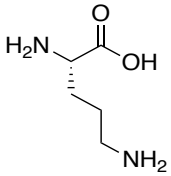
[2]Omit maps shown in blue mesh with carve=1.8. Ligands shown in sticks.

3.3.3 Comparison of P3-Variant Series

A series of pentapeptides varying at P3 (reminder P1 is the amino acid linked to the catalytic cysteine) were synthesized and their co-crystal structures with Caspase-3 were generated. The structural data would aid in comparison of conformational differences in the protein and substrate as a means of understanding potency and selectivity because of various P3 amino acids. Understanding the binding modality of the P3-variants should guide *de novo* peptide inhibitor design. Co-crystal structures were generated for 7 different P3-variant pentapeptides with a variety of amino acids used at P3: proline, valine, arginine, lysine, phenylalanine, and two unnatural amino acids: acetyl lysine and ornithine (Table 3.3). The R-groups of the P3-variants gives an interesting population for structure activity relationships as it contains charged, polar, hydrophobic, and variable extension R-groups.

Table 3.3 Structure of P3-Variants



Amino Acid	Abbreviation	Structure
Phenylalanine	F	
Lysine	K	
Arginine	R	
Valine	V	
Proline	P	
Acetyl-Lysine	Ack	
Ornithine	Orn	

3.3.3.1 Hydrogen Bonding through Peptide Backbone

In superimposing the 7 structures onto a published reference structure of Caspase-3 co-crystallized with Ac-VDVAD-CHO (PDBid: 2H65), little backbone movement is detected between the structures except for Ac-VDPVD-CHO. The pentapeptide Ac-VDPVD-CHO will be discussed in more detail in a subsequent section. The Cα's of the pentapeptides map well onto each other with some slight deviation in atomic position of the Cα of P5 (Figure 3.5). The variation of the P3 amino acid does not cause a change in the positioning of the inhibitor within the substrate binding site. The pentapeptides are able to interact through a series of conserved hydrogen bonds involving the main chain with the active site of Cas-3 (Figure 3.6). The carbonyl of P1 peptide can hydrogen bond with the amide of Cys163 and the amide of Gly122 of L1. The majority of the hydrogen bonds of the main chain of the pentapeptides involve L3. The amide of P1 hydrogen bonds with the carbonyl of Ser205. Arg207 engages the pentapeptide backbones in a variety of ways: a guanidinium nitrogen of arginine hydrogen bonds with the carbonyl of P2, the nitrogen of the arginine backbone interacts with the carbonyl of P3, and the carbonyl of the arginine main chain hydrogen bonds to the amide of P3. The side chain hydroxyl group of Ser209 can hydrogen bond with the amide and carbonyl of P5, and the amide of Ser209 also interacts with the carbonyl of P5. P4 makes only a single contact with the pentapeptide main chain: a hydrogen bond between the carbonyl of Phe250 and the nitrogen of P4. The hydrogen bond with Phe250 is not observed in the Ac-VDRVD-CHO structure, discussed in more detail below. The peptides all exhibit similar inhibition in the enzymatic activity assay which could mean that conformation of the backbone that maintains this hydrogen bonding network is potentially more important than the sequence of the amino acids within the pentapeptides. Successfully engaging Arg207 of L3 is important to developing a selective peptide inhibitor of Caspase-3.

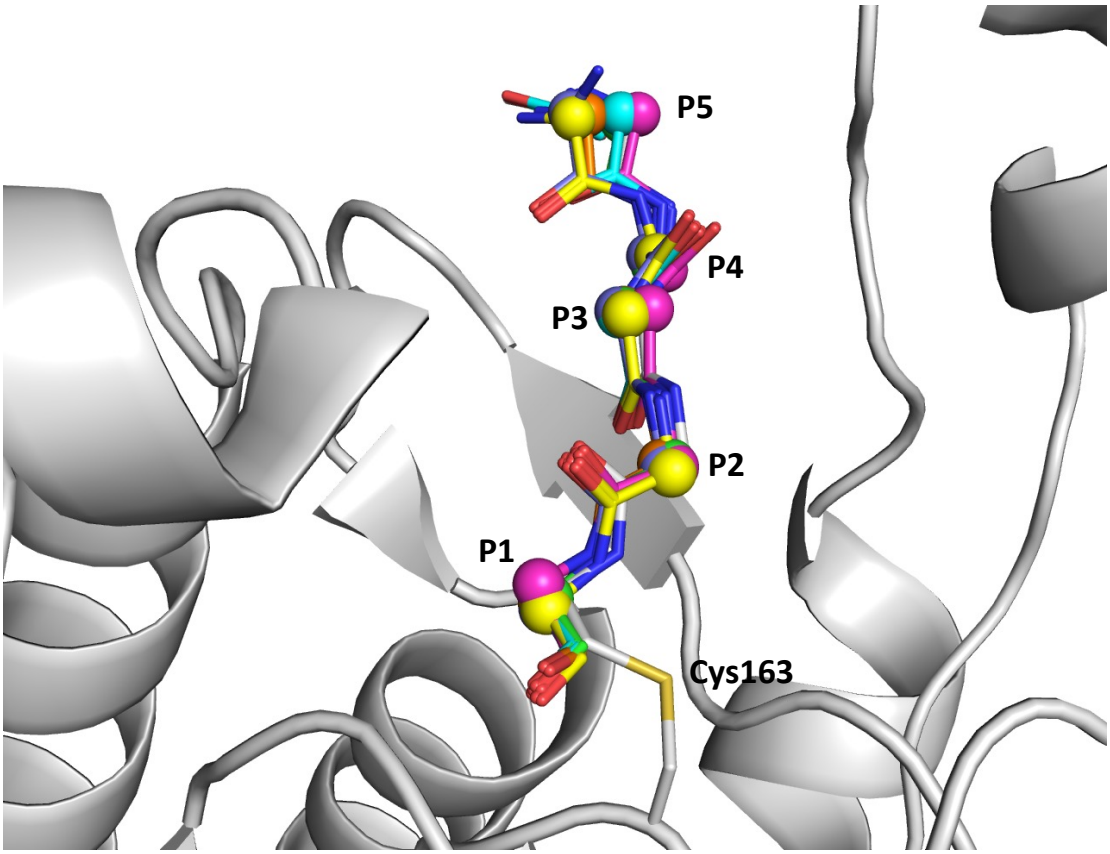


Figure 3.5 Overlay of the Backbone of the P3-Variant Pentapeptides

The main chain of the P3-variant pentapeptides (Ac-VDPVD-CHO is not shown) covalently bonded to Cys163. The structures were superimposed onto reference structure Ac-VDVAD-CHO (PDBid: 2H65). The Cα's are shown in spheres and residue numbers are labeled.

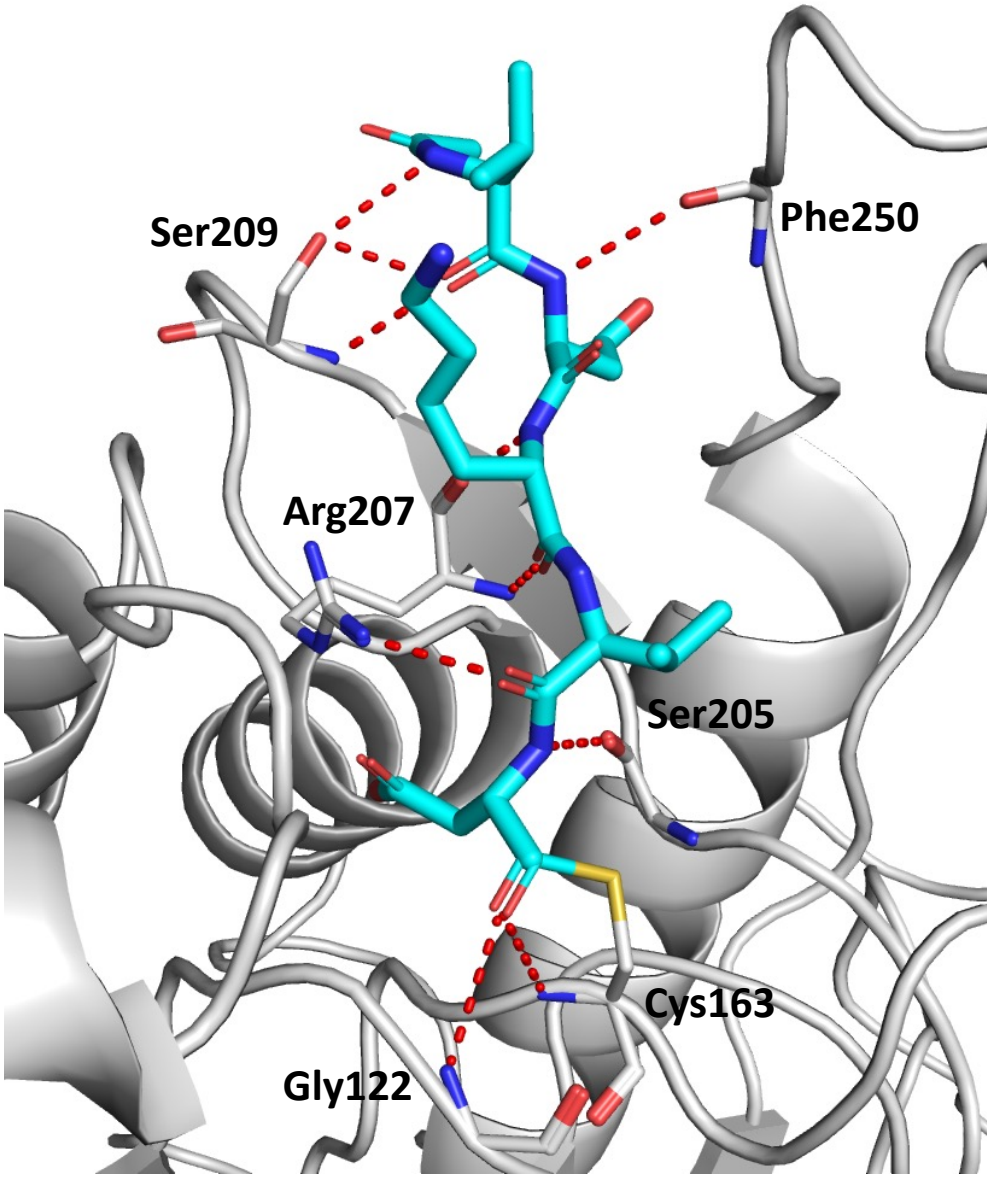


Figure 3.6 Hydrogen Bonding Network of P3-Variant Main Chain with Cas-3

3.3.3.2 L4 Loop Shows Small Shift

Looking at the L1, L2, L3, and L4 loops of the P3-variant Cas-3 co-crystal structures, little movement of the loops is detected. The molecules do not appear to cause a significant change in the tertiary structure or protein folding. The Ac-VDRVD-CHO causes a slight shift in L4 which is pushed 1.5-1.8 Å away from the substrate binding site thus increasing the exposed surface area (Figure 3.7). The movement of L4 lifts the carbonyl of Phe250 away from the amide of P4 so that

the distance between the acceptor and donor is 4.5 Å – too far for an energetically meaningful hydrogen bond with the protein (Figure 3.8). Ac-VDRVD-CHO does have slightly worse inhibition which could be explained by the loss of this hydrogen bond. The shift of L4 cannot be directly attributed to arginine at P3 as the arginine R-group does not extend towards L4 but rather lies over the pentapeptide backbone.

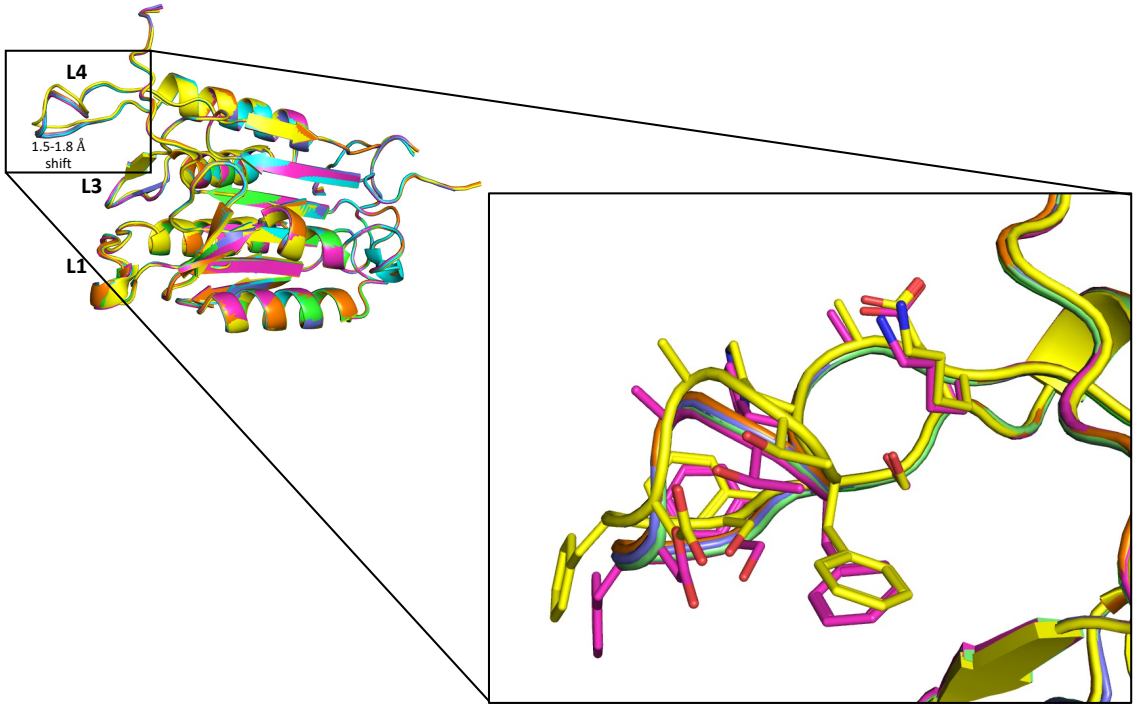


Figure 3.7 Small Shift in L4 Loop of Ac-VDRVD-CHO Structure

Overlay of the P3-variants (Ac-VDPVD-CHO is not included) with the L1, L3, and L4 loops. The L4 loop of Ac-VDRVD-CHO (in yellow) moves relative to the other structures.

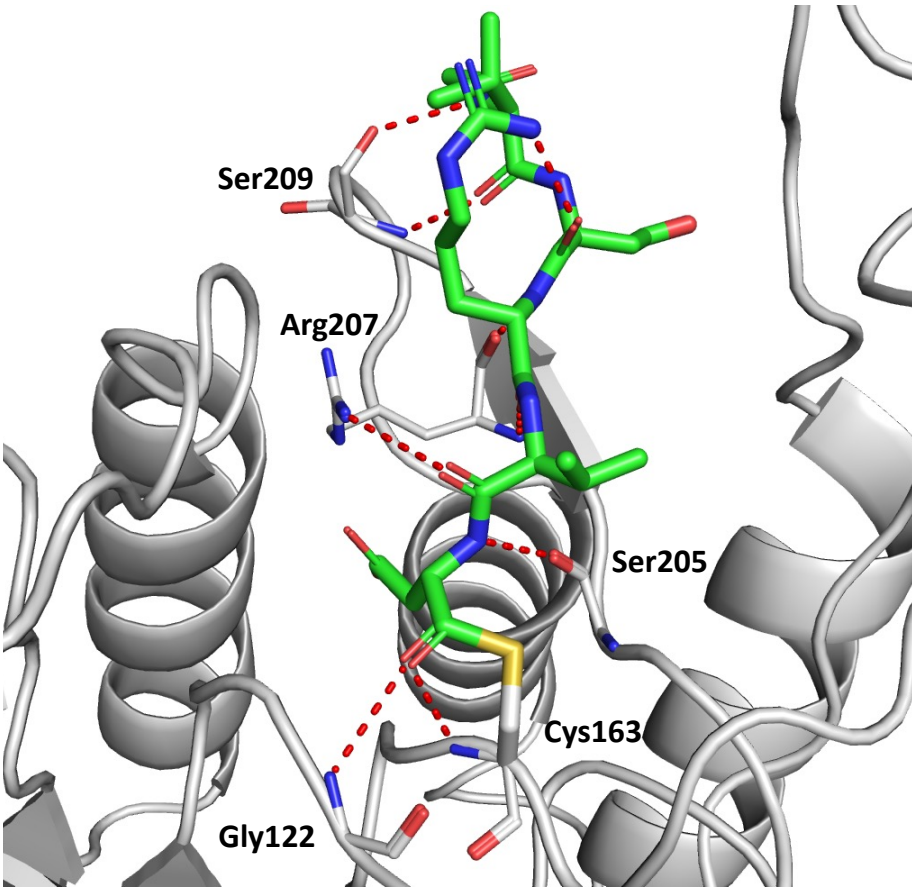


Figure 3.8 Hydrogen Bonding Network of Ac-VDRVD-CHO Backbone

3.3.3.3 Proline in Ac-VDPVD-CHO Causes Movement of P4 and P5

Looking at the K_i values we see that a majority of the modifications do not cause any change in the inhibition of Cas-3. This is not the case if a proline occupies P3. Substituting proline for a valine results in a 2-log fold reduction (34 nM vs 3519 nM) in inhibition. The pyrrolidine ring of proline restricts the phi angle and therefore causes the subsequent amino acids to be displaced (Figure 3.9). This pushes P4 and P5 (as a result of P4) specifically the aspartic acid R-group of P4 out towards the C and C α of P5 of the non-proline structures. This shift causes a loss of several hydrogen bonds of the pentapeptide backbone to the L3 loop which could explain the decrease in inhibition. Maintenance of the relative atomic positions of the pentapeptide backbone is paramount to enzymatic inhibition.

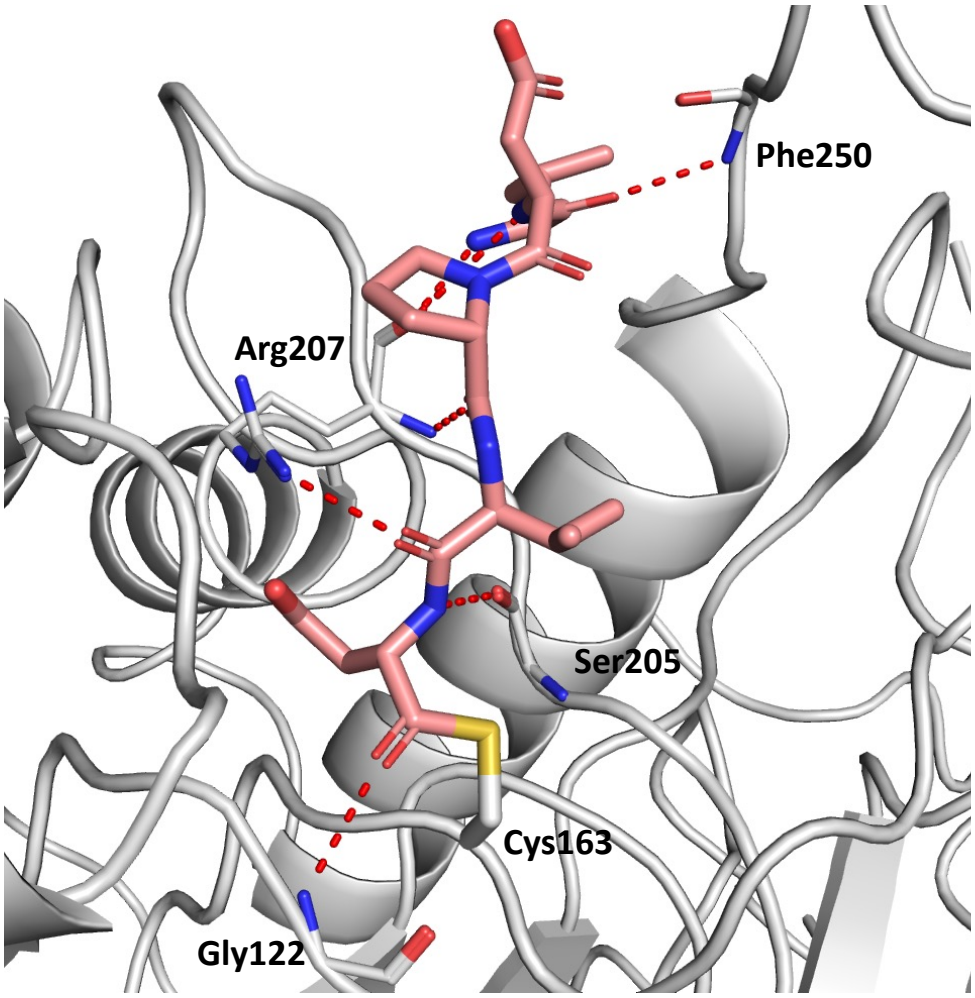


Figure 3.9 Hydrogen Bonding Network of Ac-VDPVD-CHO Backbone

3.3.3.4 Hydrogen Bonding Differences Resulting from P3

The majority of the interactions of the pentapeptide series with Caspase-3 involves hydrogen bonding of the backbone with L1, L3 and L4, as discussed in a previous section (Figure 3.6). To gain a better understanding of the structure-activity relationship of the P3 modifications, the effect of the R-group on the hydrogen bonding network was compared (Figure 3.10). Of the seven different R-groups tested, four (Ac-VDVVD-CHO, Ac-VDFVD-CHO, Ac-VDKVD-CHO, and Ac-VD(Orn)VD-CHO) are only able to interact with Phe207 of L3 through the amide and carbonyl of P3. The guanidinium nitrogen atoms of Ac-VDRVD-CHO hydrogen bond intramolecularly with the carbonyl of P4. The proline P3-variant is only able to hydrogen bond

through the carbonyl because the amide is rotated away from the carbonyl on Phe250. The only R-group that can engage any other amino acid is the acetyl-lysine P3-variant: the acetyl oxygen can form a hydrogen bond with the sidechain oxygen of Ser65. This could be a new interaction that could be targeted in new inhibitor design.

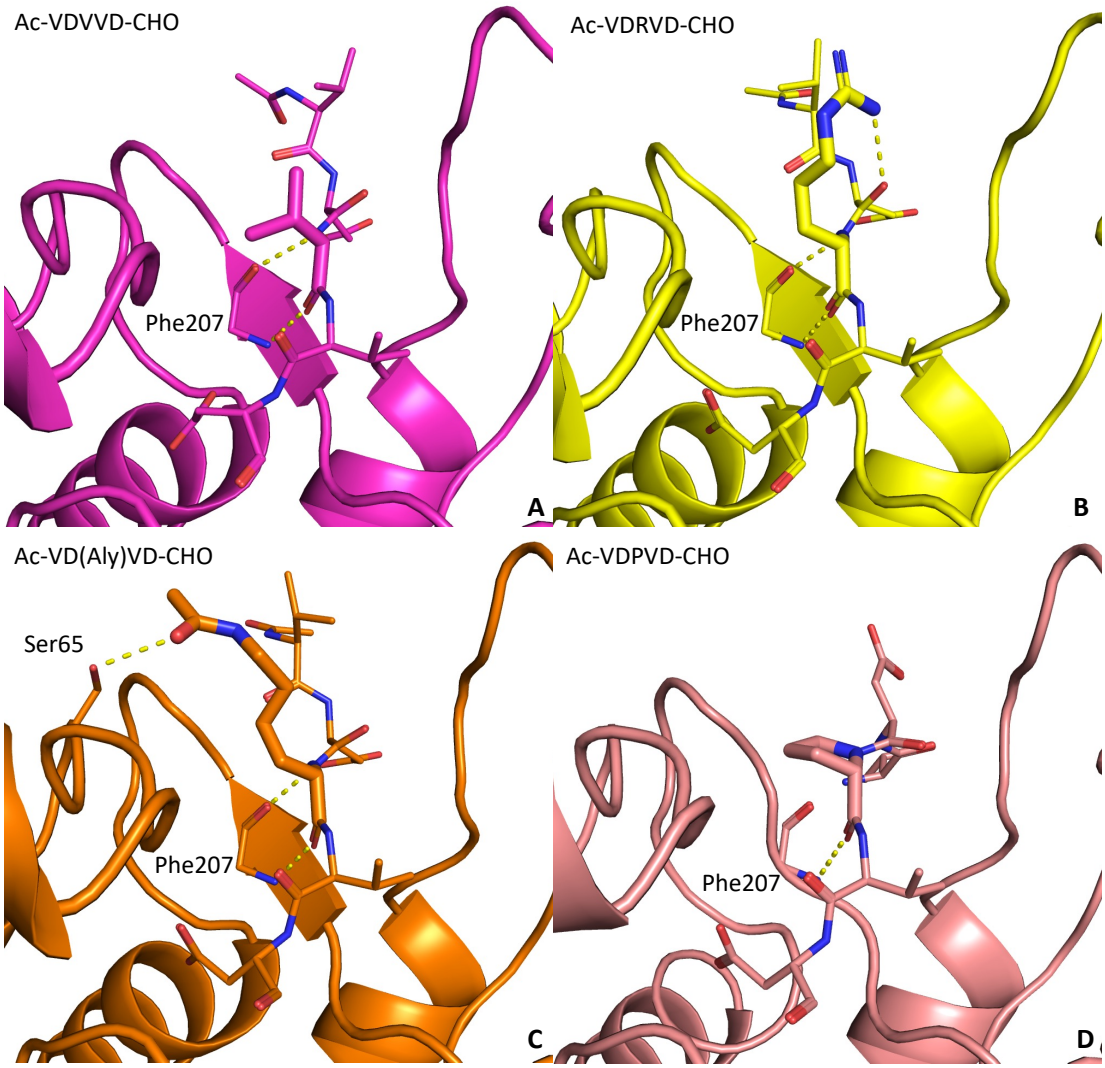


Figure 3.10 Comparison of Different Hydrogen Bonds with Different P3 Amino Acids

3.3.4 Non-canonical Peptide Structure Activity Relationship

The canonical sequence of Caspase-3 inhibitors is D-X-X-D and therefore to sample a broader amino acid profile, non-canonical peptides were co-crystallized with Caspase-3. The sequences of the non-canonical pentapeptide all contain an aspartic acid at P1 for easy covalent

bond formation with the catalytic cysteine but do not contain an aspartic acid at P4 thus making them non-canonical. The use of the non-canonical sequences is important to new peptide design as it highlights the structural role of aspartic acid at P4 on the observed inhibition. Differences in protein flexibility and inhibitor binding modality can be compared for substitutions to amino acids other than the canonical aspartic acid to guide future inhibitor design.

3.3.4.1 Ac-ITAKD-CHO vs. Ac-ITV(Dab)D-CHO

The pentapeptides Ac-ITAKD-CHO and Ac-ITV(Dab)D-CHO are very similar in their sequence but exhibit large differences in their inhibition of Cas-3 enzymatic activity. 2,4-diaminobutryic acid (Dab) is an unnatural amino acid that mimics a lysine R-group but contains one less methylene. Valine and alanine are both hydrophobic amino acids of comparable size. Therefore, the two sequences would be expected to interact very similarly with their target, Cas-3. While their atomic composition is very similar, there is a considerable difference in their inhibition on enzymatic activity (Table 3.2). This near ten-fold decrease in inhibition (moving from Ac-ITAKD-CHO to Ac-ITV(Dab)D-CHO) can be explained by comparing the co-crystal structures. The main chain of P1-P3 map well onto each other but differences are noted in the position of the P4 and P5 loop. The R-group on P2 of each pentapeptide extends towards L4 where the primary amine of lysine and Dab is within proximity of Phe256 and engages in pi-cation interactions. The Cα's of Phe250-Ala258 in the L4 loop shift 1.0-2.2 Å to accommodate the larger amino acid (lysine) at P2 exemplifying the induced-fit upon ligand binding (Figure 3.11). P4 and P5 of Ac-ITAKD-CHO follow the movement of L4 to allow for the peptide backbone to continue engaging in the hydrogen bonding network. The molecules exhibit similar hydrogen bonding networks, but Ac-ITV(Dab)D-CHO the carbonyl of P5 can engage the amide and carbonyl of Ser209 (Figure 3.12). Where positioned, the P5 nitrogen can interact with the oxygen of the backbone of Phe250. Since Ac-ITV(Dab)-D is able to make more total hydrogen bonds, it might be assumed that it would bind tighter to Cas-3 than the other non-canonical sequence, but this is not the case. The longer R-

group of lysine is causing a different induced-fit than the other ligands tested. The conformation of the L4 loop in the Ac-ITAKD-CHO may be in a lower energy conformation (protein is more stable) that is stabilized by the addition of the peptide. Generating molecules which cause this movement in L4 could increase affinity and therefore potency.

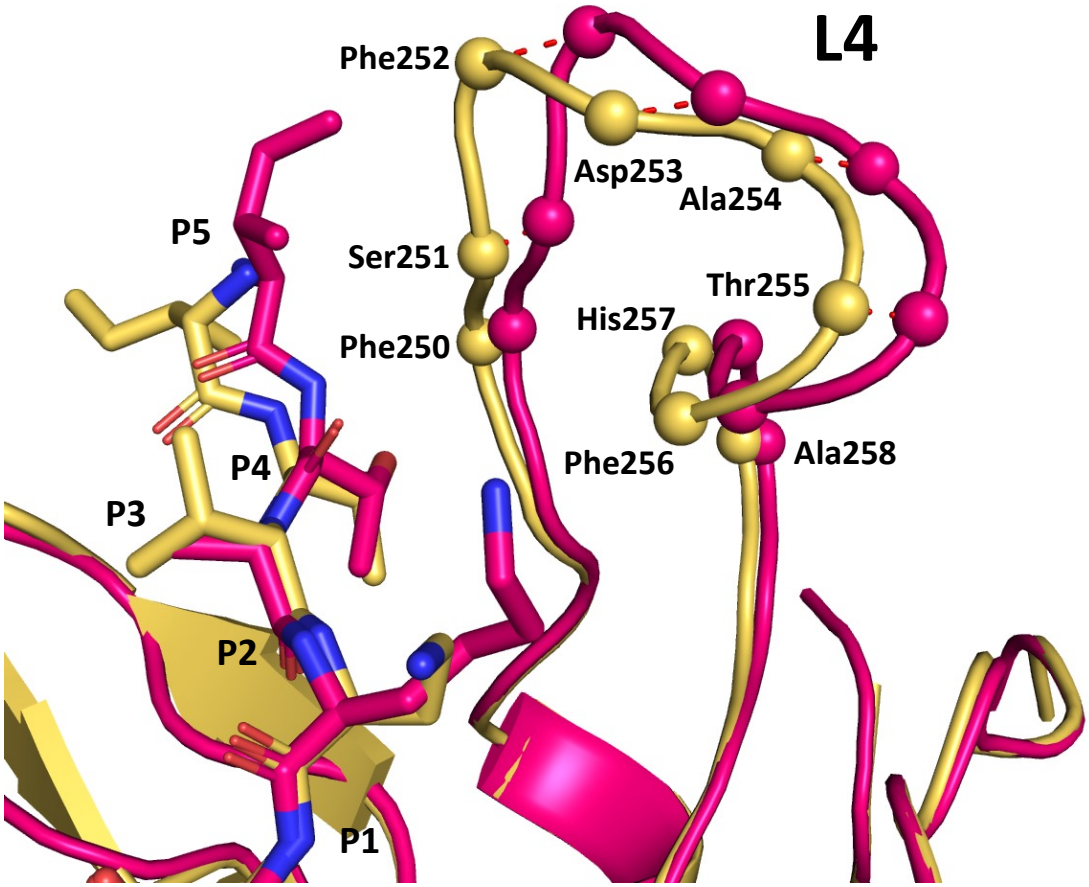


Figure 3.11 Induced-Fit by Larger P2 R-Group Causes Shift in L4

Superimposing the structures of Ac-ITAKD-CHO (fuchsia) and Ac-ITV(Dab)D-CHO (gold). The amino acids that move are shown with spheres on the Cα's.

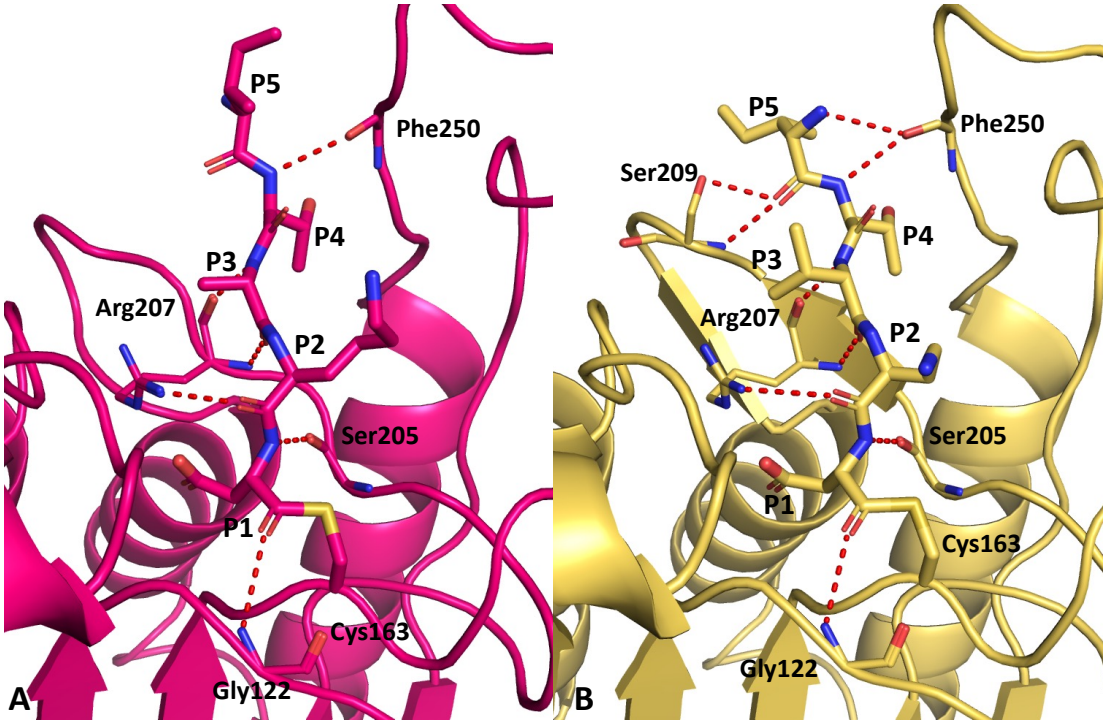


Figure 3.12 Hydrogen Bonding Comparison Between Ac-ITAKD-CHO and Ac-ITV(Dab)D-CHO

Caspase-3 hydrogen bonding with A) Ac-ITAKD-CHO and B) Ac-ITV(Dab)D-CHO

3.3.4.2 Ac-YKPVD-CHO: Absence of Density for P4 and P5

Ac-YKPVD-CHO is only modeled from P1 to P3 with nothing modeled for P4 and P5 as there was no electron density in the F_o-F_c map and modeling in the P4 and P5 peptides caused negative difference map peaks in the subsequent round of refinement. The peptide is clearly bound (Table 3.2) and the rest of the peptide would be present (no precedence for cleavage of the pentapeptide) so in crystallography this portion is deemed disordered. Crystallographic disorder describes atoms, molecules, or chains where translational periodicity, atomic positions between unit cells are related by a translational matrix, is not observed. In other words, the P4 and P5 residues adopt a variety of conformations in different unit cells and therefore atomic positions in the model cannot be defined. As discussed above with the proline P3-variant, proline causes a shift in the backbone of the protein due to the constraint on the phi angle. The pentapeptide, even with

its relative poor affinity, is able to covalently bond to Cys163. The lack of affinity highlights the importance of the hydrogen bonding network in P4 and P5 in which Ac-YKPVD-CHO is unable to engage. Design of a potent inhibitor should avoid placement of proline at P3 and maybe in the pentapeptides as a whole.

3.3.4.3 Molecular Replacement and Refinement Issues with Two Structures with Unique Cell Parameters

Generation of co-crystal structures were attempted with additional non-canonical peptides: Ac-ITV(Orn)D-CHO and ITVKD-CHO. Diffraction data was collected on both of these structures but the processing of the diffraction pattern images gave new unit cell parameters (parameters for Ac-ITV(Orn)D-CHO): 97.559, 66.67, 131.929, 90, 101.352, 90 (a, b, c, α, β, γ). Comparing these values to that of the refined structures (Table 3.1) we see that the length of the a-axis and c-axis are considerably longer. Molecular replacement using PDBid 2H65 for Ac-ITV(Orn)D-CHO placed a tetramer of the biological unit within the unit cell instead of the dimer observed in previously published structures and the other successfully generated structures in the Finzel laboratory. The additional dimer is not related by translational matrices and therefore resides in the asymmetric unit. Figure 3.2 shows that the initial morphology of the Caspase-3 co-crystals is thin plate-like crystals that can grow some greater three-dimensional character with time. The issue with viewing plate-like crystals using a microscope is determining if the plate is singular or contains multiple plate-like crystals stacked on top of each other. Twinning in crystallography refers to neighboring domains that are mutually reoriented according to a specific symmetry operation relating to the crystal lattice that does not belong to the symmetry operations relating to the space group. Twinning can usually be easily detected upon inspection of the diffraction pattern (closely related spots that vary in intensity), but current software can also be used to determine crystal twinning or presence of twin laws. Viewing the diffraction pattern and using the software present in Phenix, the poor data processing issues are not a result of twinning. The diffraction pattern has an abnormally large number of spots present on the detector which suggests that the assumed single plate-like crystal

was rather a composite of multiple plate like crystals stacked on top of each other which are not related by a crystal lattice symmetry operator (Figure 3.13). This issue was not observed in any of the other structures, and no conclusion to the overlapping plate morphology can be attributed to the ligand, the protein, or the ligand-protein complex. Decreasing the aperture of the beam and finding areas on the crystal where fewer reflections are observed (likely near the edges) should hopefully avoid this issue in the future.

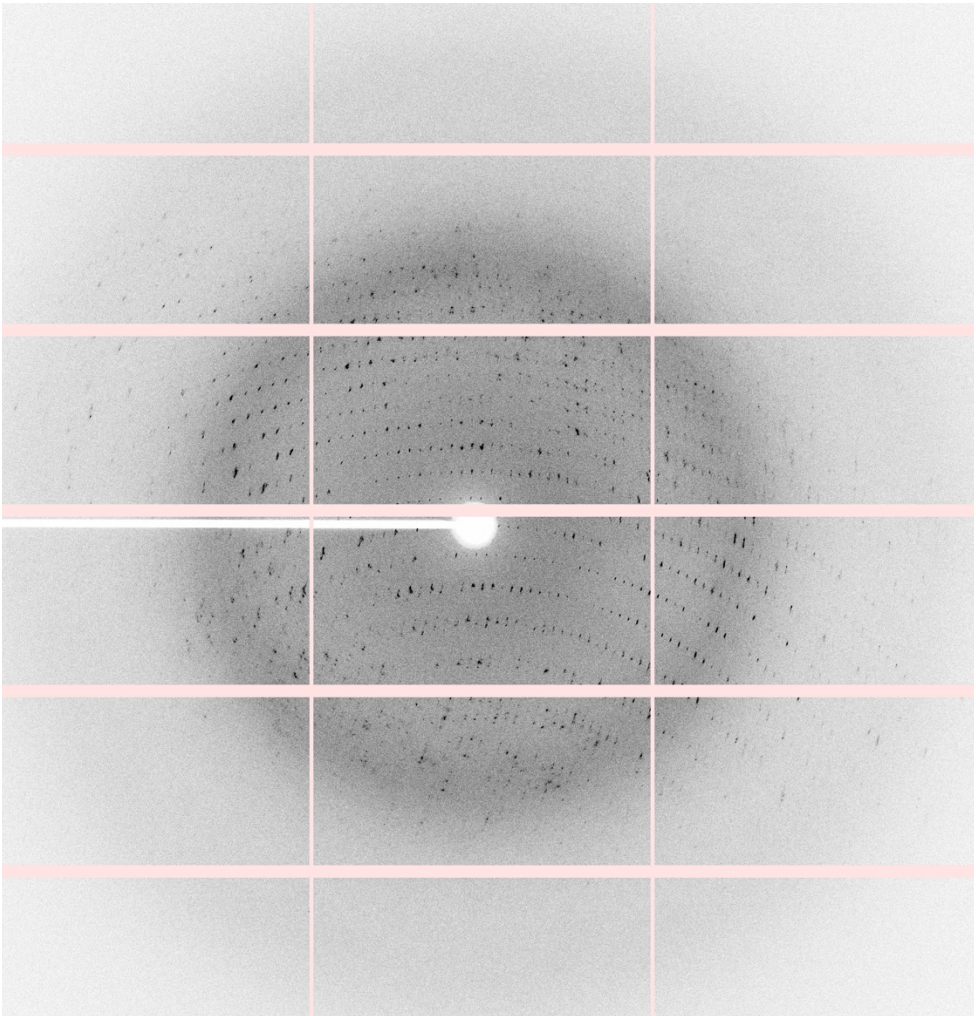


Figure 3.13 Diffraction Pattern for Overlapping Plate-Like Crystal of Caspase-3

3.3.5 Inability to Generate Co-crystal Structures with Non-Peptidic Inhibitors of Cas-3

In brief, the crystallization conditions used for the pentapeptide inhibitors were applied to a series of non-peptide like molecules. Two of the molecules were synthesized by Dr. Gurpreet Singh in Dr. Michael Walters' laboratory in the Institute for Therapeutics Discovery and Development at the University of Minnesota and have not been published, and therefore, the structure of these molecules will not be disclosed in this chapter. The synthesized molecules are non-peptide-like molecules and pose particular interest as they would have higher potential of crossing the blood-brain-brain and reaching and acting on their target, Caspase-2. The molecules using the two described crystallization conditions did not lead to crystals that diffracted (some very small micro-crystalline material was observed, but it did not grow to a size that was usable). The established method (incubation of ligand with dilute protein followed by concentrating of the protein-ligand complex) and conditions might only be conducive to peptide inhibitors. Future work on the Caspases in the Finzel laboratory will need to discover conditions that can work across all inhibitors.

3.4 Conclusion

This chapter has taken an in-depth look at the structure of Caspase-3 and the associated activity relationship of 10 pentapeptides. The pentapeptides tested covered a variety of modifications to the peptide at P3 and also examined a series of peptides that lacked the aspartic acid at P4 making them non-canonical peptide inhibitors. Examination of the P3-variant co-crystals structure shows that the pentapeptide backbone is very consistent amongst the 7 variants tested except for Ac-VDPVD-CHO. The binding of these peptides is stabilized by a series of hydrogen bonds between the main chains of the peptides and L1, L3, and L4. Arg207 makes several hydrogen bonds with the peptides and therefore engagement of this amino acid appears to be essential to inhibitory potential. The non-proline P3-variants show little change in inhibition of Caspase-3 even though the R-groups can engage in different hydrogen bonds which highlights that the R-group of

P3 is not critical to inhibition. Proline has a constrained phi angle that deviates from the angles observed in non-proline structures and this deviation causes the shift of P4 and P5 which explains the drastic loss in inhibition observed. Structural biology efforts have shown that maintenance of the hydrogen bonding network of the peptide backbone in P3, P4 and P5 is critical to the potency of the pentapeptides. Three non-canonical pentapeptides (lacking aspartic acid at P4) were co-crystallized with Caspase-3. The longer P2 R-group of Ac-ITAKD-CHO is able to push L4 away from the substrate binding site. The peptide has comparable affinity to the P3-variants and therefore this induced-fit could be forcing Cas-3 into a lower, more-energetically favored protein fold. The 10 novel pentapeptide Cas-3 co-crystal structures will hopefully be used in homology modeling and structural comparison to Cas-2 co-crystal structures leading to a potent and selective inhibitor of Cas-2.

4. Computational Efforts Towards the Design of THIQ-Saccharide Conjugates as Inhibitors of CD44

Portions of this work have excerpted from published work.

Table 4.1 was remade with permission from: Li-Kai, L. Discovery of Small Molecule Inhibitors of Hyaluronan Binding at Cell Surface Receptor CD44. *Thesis Dissertation*. **June 2015**

Author Contributions: Dr. Soma Maitra in Dr. Gunda Georg's laboratory at the University of Minnesota completed the synthesis and characterization of the THIQ-Monosaccharide conjugate. Scheme 4.1 depicts the synthetic scheme to generate the conjugate molecule and was made by Dr. Soma Maitra in ChemDraw. The synthesis is not described in any detail.

A detailed background on the role of the extracellular matrix in cancer, the function of the signaling molecule HA, a comparison of the hyaldherins, an in-depth look at the known structural biology surrounding CD44 and its interaction with HA, and a look at antagonism of CD44-HA interactions is described in Chapter 1.

4.1 Introduction

4.1.1 Rational Behind THIQ-Oligosaccharide Conjugate Molecules

Previous biophysical and structural biology studies in the Finzel lab resulted in the identification of a series of small molecules containing a tetrahydroisoquinoline (THIQ) moiety that bind in an inducible small-molecule subsite of CD44 HABD with sub-millimolar affinity.[88] A Maybridge Ro3 Diversity Fragment library of 1000 compounds was screened in a single point immobilized-HABD surface plasmon resonance (SPR) assay with a cutoff threshold of approximately 10 RU to achieve an initial hit rate of 4%.[88] Hit fragments were then screened in a dose-response SPR assay to find compounds that could be fitted using a 1:1 binding isotherm.[88] Molecules that fit this binding isotherm were then subjugated to co-crystallization with *m*CD44 and structures were generated for five of these molecules.[88] Two of the small molecules, an aniline-based compound and a THIQ-based compound, were of particular interest for their binding location and induced conformational changes in the protein (Table 4.1). The addition of the exocyclic

nitrogen present in the aniline-based compound to the THIQ moiety led to a 7-fold improvement in binding affinity.[88] In comparison of the *apo* structure to the small-molecule induced structure, Arg155 and Asn29 rotate to create pi-stacking above and below the aromatic ring while Glu41 moves away from Arg82 to form an H-bond with the THIQ secondary amine and to form a hydrogen bond with the reoriented Asn29 (Figure 4.1). The exocyclic nitrogen on the THIQ moiety can form a hydrogen bond with the backbone oxygen of Val30 and two well-anchored water molecules (Figure 4.1). Crystal structures of a 5-amino-THIQ and 8-amino-THIQ shows some flexibility in the orientation of the THIQ. The endocyclic nitrogen in both structures can hydrogen bond to Glu41 (alternate Glu41 rotamers) by adopting different ring puckers (Figure 4.1). An 8-amino-THIQ scaffold was selected over the initial 5-amino-THIQ hit compound as the 5-amino compound introduces a chiral center which decreases the synthetic opportunities and the 8-amino-THIQ lies in closer proximity to the Glc-5 subsite.[227]

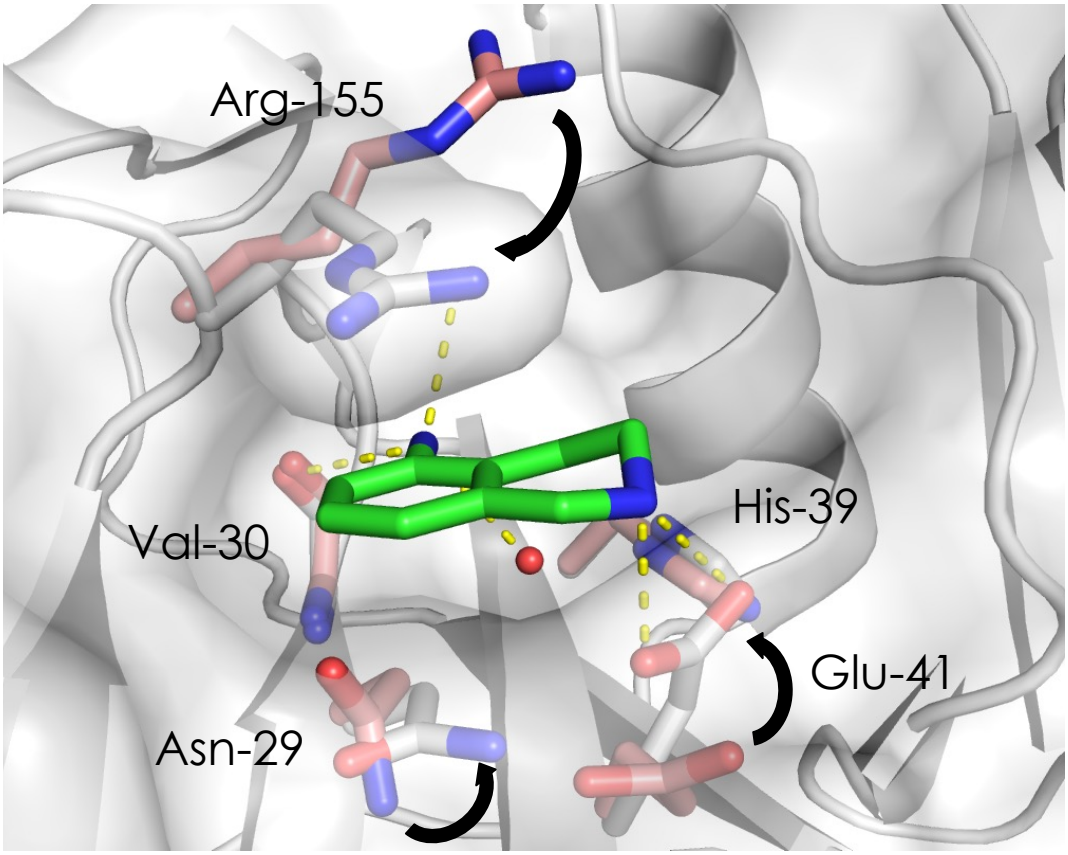
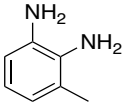
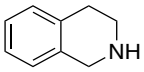
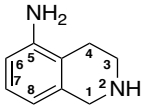
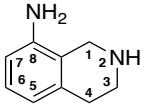
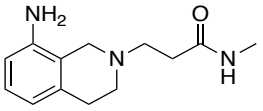
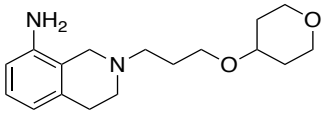


Figure 4.1 Conformational Changes in *m*CD44 Induced by THIQ

5-amino-THIQ induces conformational changes in Arg155, Asn29 and Glu41 to bind in a site just below the HA binding groove in CD44. The apo structure is colored in salmon (PDBid: 4mrd) and the bound structure is colored in white (PDBid: 5mrg). Conformational differences are highlighted with arrows.

Table 4.1 Fragment Evolution of THIQ Moiety

Structure[1]	LE[2]	K_D (mM)[3]		IC_{50} (mM)[3]	PDBid:
		h HABD	*m* HABD	*h* HABD	
	0.31	8.5 ± 1.2	6.4 ± 1.0	No Inhibition	4MRE
	0.29	6.9 ± 1.5	11.2 ± 4.5	27	4MRF
	0.4	0.9 ± 0.2	1.1 ± 0.2	6.7	4MRG
	0.36	1.2 ± 0.2	No Data	7.5	5BZI
	0.26	0.6 ± 0.4	No Data	3.1	5BZO
	0.2	0.9 ± 0.2	No Data	3.4	5BZR

[1]The numbering of atom position is provided for the 5-amino- and 8-amino-THIQ compounds to help clarify the difference between the analogs. The endocyclic nitrogen of the isoquinoline is atom 2.

[2]Ligand efficiency, (–RT ln*KD*)/(number of non-hydrogen atoms), was calculated based on the affinity of ligands to *h*HABD in *kcal mol*$^{-1}$.

[3]KD and IC50 values were measure using the two surface plasmon resonance assays: the immobilized HABD SPR assay and the immobilized HA SPR assay, respectively.

A series of 8-amino compounds were synthesized with extending length from N2 and were found to exhibit increasing potency and inhibition with further extension towards the HA groove (Table 4.1).[227] A *N*-methylpropanamide extension from N2 of 8-amino THIQ was purchased from Enamine. The three carbon linker extends towards and into the hyaluronan binding groove with the nitrogen lying in almost the same position as the 3' hydroxyl of the Glc-5 of HA in the co-crystal structure of *m*CD44 with an HA-oligosaccharide, PDBid: 2JCR (Figure 4.2) As a proof of concept,

a pyran compound with a three-methylene and one oxygen atom linker connected to N2 of the 8-amino-THIQ was synthesized in the ITDD at the University of Minnesota.[227] The compound allows for proper positioning of the THIQ moiety and the pyran reaches into the Glc-5 subsite but does not superimpose directly onto Glc-5.[227] The findings suggest that the three-carbon linker leading into Glc-5 should allow for proper conformations of the THIQ and Glc moieties to allow for improved binding affinity.

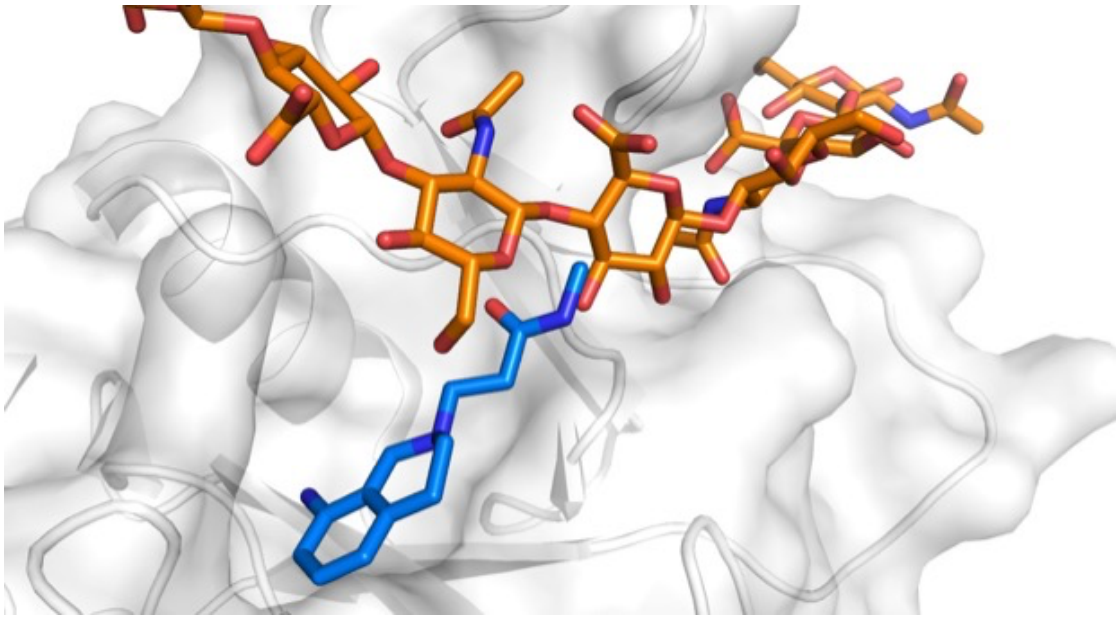


Figure 4.2 Overlay of THIQ-Methylpropanamide and HA-Oligosaccharide Structures

Overlay of co-crystal structures of CD44 (white cartoon and white surface) with THIQ-N-methylpropanamide conjugate in blue (PDBid: 5BZO) and HA-oligosaccharide in orange (PDBid: 2jcr).

4.1.2 Fragment Linking to Increase Potency

In the drug discovery process, one approach to improving the potency of fragment molecules involves linking two fragment molecules to form a larger non-fragment like molecule.[228] The approach is contingent on discovery of multiple hit fragments whose structure can be determined using structural biology approaches and their binding modality places the fragments within different but closely located sub-sites within the protein binding site.[228] Linking of two

fragments should be greater than the sum of the binding energies of the two fragments and instead a "superadditivity" or "positive cooperativity" should be observed.[228,229] This principle is described by Jencks in Equation 4.1.[229] The Gibbs free energy of binding of the linked fragments is quantified as the free energy of each individual fragment with the addition of ΔG_S which takes into account changes in translational and rotational entropy as a result of the introduction of the linker.[229,230] Equation 4.1 can be rewritten to describe the dissociation constant K_D^{AB} for a linked fragment A-B in Equation 4.2.[231,232] The dissociation constant is the product of the two individual dissociation constants and a factor termed linking coefficient, E.[231,232] An ideal linking coefficient would be less than 1 and less than optimal or linker resulting in negative cooperativity would give an E value greater than 1.[232] The THIQ-Monosaccharide was designed with a three-methylene linker in hopes of generating a molecule with superadditivity to generate a more potent inhibitor with selectivity for CD44 over the other hyaladherins (Figure 4.3). To test the linker design, computational efforts in the form of docking and scoring were employed and the conjugate molecule was found to have greater computational affinity than either of the individual components. The conjugated THIQ molecule was then tested in an SPR assay and was found to not inhibit HA-CD44 interaction. A CD44-THIQ-Monosaccharide conjugate co-crystal structure could not be solved using soaking techniques which had proved successful for other CD44 structures. To understand the lack of binding affinity observed, a conformational search *in silico* was performed and found that a greater than 6 kcal/mol potential energy difference existed between the ideal binding pose as modeled and the lowest energy conformation. Docking and scoring efforts were used again to assess the computational binding affinity and deviation from ideal binding pose of the THIQ-Disaccharide conjugate and 8 related mimetics. Our findings suggest that the 4'-hydroxyl is critical to the observed affinity and deletion of the 2'-hydroxyl might impart enough conformational flexibility to allow for proper binding of the THIQ and saccharide moieties. While we did not get any biophysical or structural biology data on the THIQ-Monosaccharide conjugate, we suggest design

and synthesis of the THIQ-GCU-NAG conjugate and the THIQ-d2-OH-GCU-NAG conjugate for their improved binding affinity as potential inhibitors of CD44.

Equation 4.1 Gibbs Free Energy of Binding Derived From Linking Two Molecules

$$\Delta G_{AB} = \Delta G_A + \Delta G_B + \Delta G_S$$

Equation 4.2 Dissociation Constant for Linked Molecule

$$K_D{}^{AB} = K_D{}^{A} K_D{}^{B} E$$

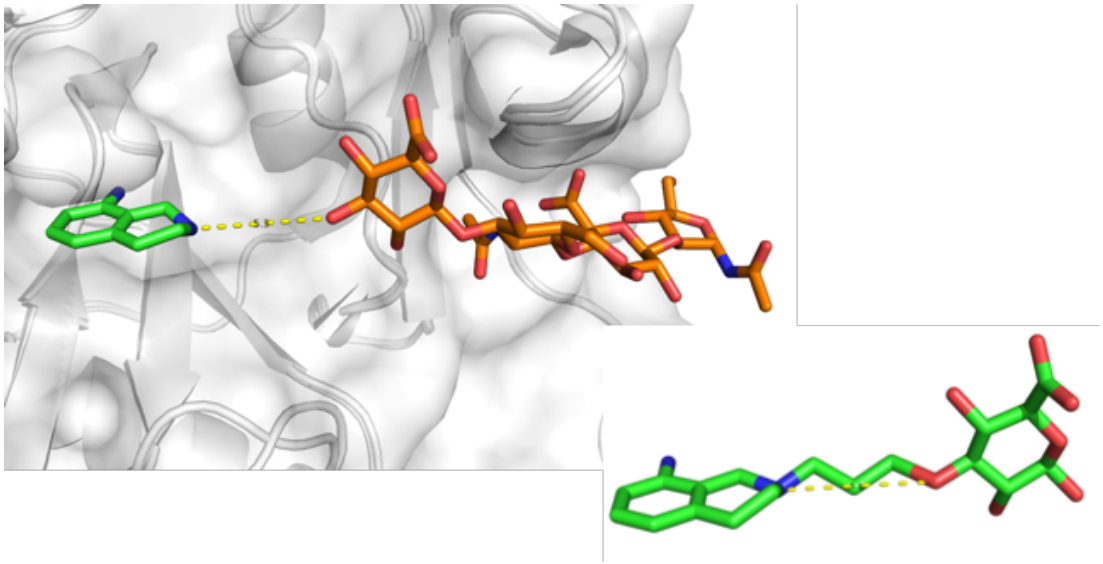


Figure 4.3 THIQ-Linked Oligosaccharide Development

Overlay of 8-amino THIQ (PDBid: 5BZI) with HA-tetrasaccharide (PDBid: 4MRD) both bound to *m*CD44. Interatomic distances between the endocyclic nitrogen of the THIQ and the 3' hydroxyl of Glc-5 is 5.3 Ångstroms and the proposed THIQ-monosaccharide conjugate shown in the lower left having a distance of 5.1 Å.

4.2 Methods

4.2.1 Reagents

All computational studies were completed using *Maestro*[233] part of the *Schrodinger Software Suite* hosted through the Minnesota Supercomputing Institute (MSI) at the University of Minnesota. Full-length human CD44 (*h*CD44) cDNA and murine CD44 (*m*CD44) cDNA were a

gift of Dr. James B. McCarthy at the University of Minnesota. Rosetta 2(DE3)pLysS cells were purchased from EMD Biosciences (Gibbstown, NJ) and competent cells were made by using a Mix and Go! E. coli Transformation Kit from Zymo Research (Irvine, CA). HiPrep 16/60 Sephacryl S-100 HR was obtained from GE Healthcare (Piscataway, NJ). High molecular weight hyaluronan (HMW-HA) from rooster comb in the form of sodium salt was purchased from Sigma-Aldrich (St. Louis, MO). CM5 and SA Biacore Sensor Chips were acquired from formerly GE Healthcare (Piscataway, NJ) now Cytiva. L-Arginine, reduced/oxidized glutathione, coupling reagents, and other general buffer reagents were from Sigma-Aldrich (St. Louis, MO). The hydrazide biotin linker for biotinylated HA was from Thermo Scientific (Rockford, IL). The crystallography supplies including 24 well plates, coverslips and crystal-crushing tools were obtained from Hampton Research (Aliso Viejo, CA).

4.2.2 *Protein Preparation Wizard*[234] in Schrodinger Suite

The structure of murine CD44 co-crystallized with a tetrahydropyran linked THIQ, PDBid: 5bzr[227], was used for the computational studies as it contained the pharmacophore of interest in the small-molecule binding pocket and extends into the hyaluronan binding groove. The PDB file was imported into *Maestro*[233] and the *Protein Preparation Wizard*[234] was employed to ready the model for subsequent computational docking. Missing hydrogen atoms were added, co-crystallized waters were removed, and missing atoms and residues were added prior to running *Prime*[235], Schrodinger's protein structure predictive program. Finally, a restrained minimization of the structure was completed to sample hydrogen atom position and relax strained bond, angles, and clashes.

4.2.3 Ligand Preparation Using *LigPrep*[236]

The mono-, di-, tri-, and tetra-saccharide HA molecules were generated from the murine complex of CD44 with an HA 4-mer (PDBid: 4mrd) by extracting the saccharide portion and saving as the desired lengths. The following ligand preparation was not done on the HA-oligosaccharides as the molecules were imported and scored in place instead of a traditional docking and scoring

approached detailed below. The THIQ-modified oligosaccharides were made in JLigand[237] and atomic positions superimposed onto the 8-amino-THIQ in PDBid: 5bzr[227] and the saccharide of PDBid: 4mrd[88]. The disaccharide mimetics were generated by deleting specific atoms of the previously generated THIQ linked disaccharide molecule. Ligand structures were imported into *Maestro* as .mol2 files and *LigPrep* was employed to properly prepare ligands for docking in the computational studies. The *OPLS3* force field was selected for minimization.[238] Possible ionization states at a target pH of 7.0 ± 2.0 (to mimic physiological pH) were generated using *Epik*.[239] Tautomers of the structures were not generated for any molecules. Specified chiralities were retained and the number of poses per structure were not limited.

4.2.4 Computational Docking using *Glide*[240] in *Maestro*

Docking in *Maestro* using *Glide*[240] employs a grid for where docked ligands must bind, but to further insure proper atomic positions of the THIQ-moiety relative to previously published structures several additional constraints were implemented. To begin, a grid was established that was a cubic area surrounding the small molecule in the prepared protein structure. To ensure proper docking of the THIQ moiety and the linker leading to the saccharide moiety, positional constraints were placed on the endocyclic and exocyclic nitrogen atoms of the THIQ and on the oxygen of the linker restraining the atomic positions of the docked molecules to within 1 Å of the reference pose. Hydrogen bond acceptors/donors were specified for Val30, Glu41, and His39 of the receptor protein based off the previously known hydrogen bonding network. Excluded volume spheres were placed in the HABD groove where NaG-4 of the HA 8-mer would bind to aid in placement of Glc-5 in the groove. A core pattern comparison was used to restrict docking position of the heavy atoms of THIQ and the linker with a tolerance of 0.1 Ångstroms. Torsional constraints were placed on the carbon linkers of the molecules. For docking of the unmodified HA-oligosaccharides, the additional constraints were not used as most of the constraints do not apply to oligosaccharide binding and docking could easily replicate atomic coordinates detailed before without the additional constraints.

Van der Waals radii were left as their default setting with a scaling factor of 0.8 and partial charge cutoff of 0.15. Extra precision (XP) was used with flexible ligand sampling including nitrogen inversions and ring conformations. *Epik*[239] state penalties and large excluded volume penalties were included to force poses to be closer to expected binding poses. Docking poses were compared to the proposed/expected binding pose using RMSD values of heavy atoms in the ligands of interest. For more flexible docking of the simplified glucuronic acid disaccharide conjugates, positional and torsional constraints and excluded volumes were removed from the generated grid to further sample conformational space of target molecules.

4.2.5 Conformational Search Using Molecular Mechanics

A conformational search of the THIQ-saccharide conjugates was performed using the *Conformational Search*[241] in *Maestro*. The THIQ pharmacophore was kept fixed and all other atoms and angles were allowed to freely move and rotate. An *OPLS3*[238] force field was applied for the molecular mechanics calculation and conformations were ranked by their relative potential energies. A total of 50 conformational states were generated for each ligand. To compare the conformations to the expected binding conformation, rms calculations (atom criteria was changed depending on the specific target of the search) were used to find the ideal pose from the various conformations.

4.2.6 Cloning, Expression, and Purification of Human CD44 (*h*CD44) HABD and Murine CD44 (*m*CD44) HABD

Soluble and properly folded hyaluronon binding domain of human CD44 (*h*CD44 HABD) was expressed and purified as previously described.[88] In short, $h\text{CD44}_{20\text{-}178}$ was transformed into Rosetta2(DE3)pLyss cells and plated on LB Amp100 Cm30 agar plates. A single colony was selected from the overnight incubation and added to a vial containing 5mL of LB Amp100 Cm30 and allowed to shake overnight at 270 rpm and 37°C. 1L LB media was inoculated with 3 mL of the overnight culture and placed in the incubator using the same parameters until mid-log phase

(OD_{600} = 0.6-0.8) was reached. The culture was then rapidly chilled in an ice bath until the temperature of the culture was below room temperature prior to induction with 0.4 mM isopropyl-β-D-thiogalactosidase (IPTG). The culture was induced for 4 hours at 30°C and 270 rpm and then the cells were harvested by centrifugation at 5000*g* for 15 minutes. Resulting cell pellet was stored overnight at -20°C. Protein isolation and refolding was completed using previously described methods.[86] The inclusion bodies were isolated using iterative rounds of washing with TritonX-100 and use of a Dounce Homogenizer. The pellet was then solubilized at 4°C in an 8 M urea buffer overnight. The unfolded protein was added to a refolding buffer containing 250 mM arginine, 2 mM reduced glutathione, and 1 mM oxidized glutathione using an Econo Gradient Pump for rapid dilution to reduce aggregation and allow for proper refolding. The protein was allowed to stir in the refolding overnight at 4°C prior to concentrating the protein using an Amicon Stirred-Ultrafiltration cell to ~5mL. Monomeric *h*CD44 HABD was isolated from oligomeric forms of the protein using a HiPrep 16/60 Sephacryl S-100 column equilibrated with SEC Buffer (20 mM Tris pH 8.0 and 150 mM NaCl) and fractions analyzed on an SDS/PAGE gel. To the fractions corresponding to monomeric HABD was added 10% glycerol and then flash-frozen before storage at -80°C for future use.

4.2.7 Immobilized-HA Surface Plasmon Resonance Assay

To confirm the binding affinity a previously published biophysical assay, an immobilized-HA surface plasmon resonance (SPR) assay, was employed.[88] In brief, high molecular weight-HA (HMW-HA) was biotinylated using hydrazide biotin as previously described,[242] and was immobilized onto a streptavidin SA Biacore Sensor Chip using established methods.[88] Biotinylated-HA in a HEPES-EP Buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% (v/v) Surfactant P-20) was passed over the streptavidin chip surface until the response signal stabilized at 115 RU as monitored with the Biacore S200. The chip surface integrity and the competition-based assay were confirmed using a positive control, HA-10, in a two-fold dilution

series incubated with a fixed concentration of *h*CD44 (~17 mM) on a plate shaker for 10 minutes at room temperature to allow for binding of the small molecule to *h*CD44. Samples in a 96-well format were placed on the Biacore S200 and injected over the chip surface at 30 μL/min for 1 minute followed by 2 minutes of a dissociation phase using the same HEPES-EP buffer. Following the confirmation of the IC_{50} value for HA-10, the assay was applied to the THIQ-monosaccharide conjugate solubilized in DMSO. The same SPR experimental parameters used for HA-10 were employed. Values taken 5 seconds prior to the end of the association phase were used in subsequent analysis.

4.2.8 Attempted Co-crystallization of *m*CD44 with THIQ-Monosaccharide Conjugate

Crystallization of *m*CD44 was reproduced using previously published conditions.[85,88] *m*CD44 frozen at -80°C was thawed on ice prior to concentrating to 8 mg/mL. Protein was mixed in a 1:1 ratio with well solution containing 30% PEG MME 5000, 200 mM Na_2SO_4 and 100 mM MES pH 6.5. Monoclinic form crystals grew in about 4 days as polymorphic plates and were used in subsequent rounds of seeding to yield separate plate-like crystals. *Apo m*CD44 HABD crystals were transferred to a drop containing well solution supplemented with 10% glycerol and THIQ-monosaccharide conjugate at 10 mM. Crystals were allowed to soak for a variable period from 5 minutes to 1 hour prior to being flash frozen in liquid nitrogen and ready for data collection.

4.2.9 X-Ray Data Collection and Structural Determination

X-ray diffraction data was collected at IMCA-CAT beamline 17-ID at the Advanced Photon Source (APS), Argonne, Illinois, USA. Collection was conducted at 100 K using 1.00 Ångstroms wavelength radiation on a Dectris Eiger2 9M detector. Data was processed using *autoProc* and *aP_scale* using R-factor (<.4), completeness (>.9) and I/sigma (>2) as criteria, with a minimum of two of the three criteria being met to determine the proper resolution range.[201] The *m*CD44 model with a bound THIQ-tetrahydropyran molecule (PDBid: 5bzr[227]) was used as the

search model in molecular replacement with *Phaser*.[202] The monoclinic crystal in space group $P2_1$ was found to contain two protein chains in the asymmetric unit, as previously described, and iterative rounds of refinement and model building were carried out using *Phenix*[203] and *Coot*[204,243]. The THIQ-Monosaccharide conjugate was not found in any of the structures solved and therefore coordinates and associated reflection data are not included in this chapter.

4.3 Results and Discussion

4.3.1 Increasing Length of HA-Oligosaccharides Exhibit Improved Docking Scores

Previous biophysical efforts have shown that increasing length of HA-oligosaccharides (HA-4 to HA-12) exhibit greater binding affinity and inhibitory potential.[88] The smallest HA-oligosaccharide that could be quantified in the two SPR assays was HA-4.[88] Our initial computational efforts sought to quantify the binding affinity of the HA-oligosaccharides as a means of better understanding the binding modality of each increasing saccharide unit and to replicate the poses of previously published structures.[85,88] We focused on the HA-oligosaccharides binding between Glc-5 and NaG-8 in the hyaluronan binding domain groove as our THIQ-saccharide conjugates extend towards this portion of the groove. Our docking studies were able to replicate the hydrogen bonding network of previous crystallographic data which proves that our computational methods are an accurate measurement tool given the current model (Figure 4.4) Binding of the docked tetrasaccharide is stabilized through hydrogen bonds with 7 different amino acids within the HA-binding groove. The NaG-8 is slightly rotated relative to the crystal structure of *m*CD44 with the tetrasaccharide but is still able to engage in the hydrogen bond. Docking scores increase (a more negative number – further away from 0) with each increasing saccharide unit added. We see the largest relative change in docking scores moving from the monosaccharide to the disaccharide (-2.886 vs. -4.699) and little increase moving from the trisaccharide to the tetrasaccharide (-5.849 vs. -5.965), (Table 4.2). It has been well established that HA makes weak interactions within the binding groove but that NaG-6, through the *N*-acetyl functional group, is

able to form hydrogen bonds with Ile100 and Tyr46 which explains the large increase in docking scores moving from the monosaccharide to the disaccharide. NaG-8 is able to make one additional hydrogen bond with Tyr109 but the entropic cost of binding could explain the relatively small improvement in docking scores. Further inhibitor design should attempt to target the sites occupied by NaG-6 and Glc-7 as these sites play a key role in the affinity of hyaluronan for the binding groove.

Table 4.2 Docking Scores of HA-Oligosaccharides and THIQ-Saccharide Conjugates

Compound	**Docking Score**
Monosaccharide (Glc-5)	-2.886
Disaccharide	-4.699
Trisaccahride	-5.849
Tetrasaccharide	-5.965
THIQ-Monosaccharide Conjugate	-6.441
THIQ-Disaccharide Conjugate	-8.083
THIQ-Trisaccharide Conjugate	-10.076

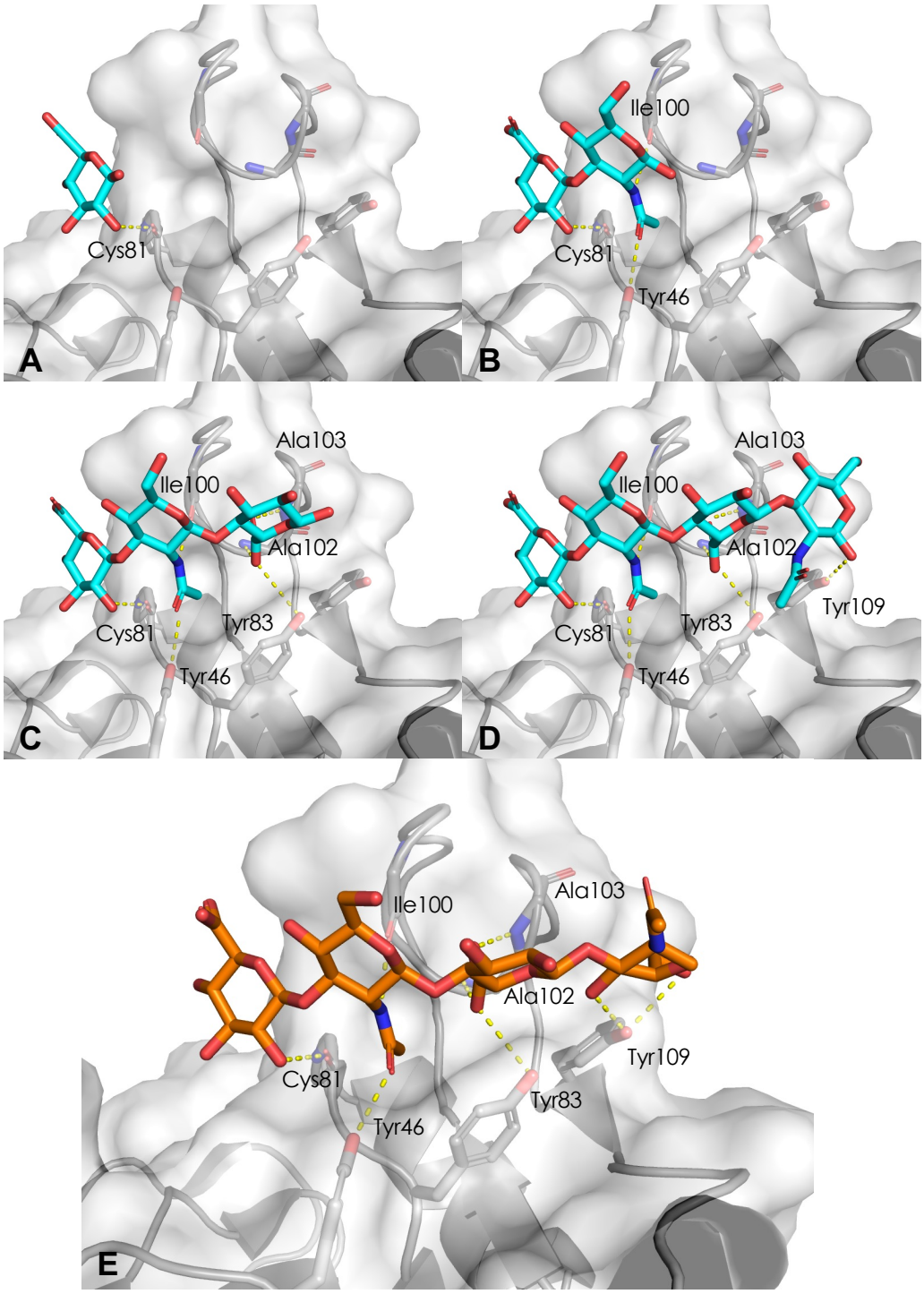


Figure 4.4 Hydrogen Bond Comparison Between Docking and Published Structures

A) Monosaccharide, B) Disaccharide, C) Trisaccharide, and D) Tetrasaccharide in cyan as the preferred binding pose. E) Tetrasaccharide in published structure (PDBid: 4mrd) in orange.

4.3.2 Docking Studies Suggest that THIQ-Saccharide Conjugates will have Higher Binding Affinity

Following the determination that docking scores improved with increasing length of HA-oligosaccharides, the THIQ-saccharide conjugates were tested. Initially, the ligands were tested using the same parameters for docking in *Glide,* but the poses generated were deemed poor as the location of the THIQ pharmacophore was not found in the small-molecule subsite of *m*CD44 and was not making any of the known key interactions. Therefore, to force the docking poses to resemble the proposed binding pose more closely, by keeping the positioning of the THIQ and saccharide moiety intact, positional, torsional, and excluded volume constraints were imposed (Figure 4.5). With the constraints in place, the number of plausible binding modalities was decreased and therefore only a select number of docked poses had to be analyzed. The small number of resulting poses were compared to the proposed conformation using rms calculations. The docking scores of the preferred binding poses were analyzed and as the length of the THIQ-saccharide conjugate is increased there is a proportional increase in the docking scores (Table 4.2). While the docking of the saccharide molecules and the THIQ-saccharide molecules use different constraints to obtain the docking poses and scores, the scores can be compared as the same force field, *OPLS3* is applied to all the molecules. The docking scores for the conjugates are all greater (more negative) than the individual saccharide molecules and therefore conjugation of the two molecules creates tighter binding as expected by linking two fragment molecules. The molecules are able to branch the two binding subsites and allow for binding of both the THIQ-pharmacophore and proper placement of the saccharide unit within the hyaluronan binding domain (Figure 4.6). This supports the three-methylene linker design linking N2 of 8-amino-THIQ to the 3'-hydroxyl of Glc-5. The docking studies suggest that synthesis of the disaccharide and trisaccharide molecules should afford tighter binding and therefore greater binding affinity and potentially higher inhibitory potential.

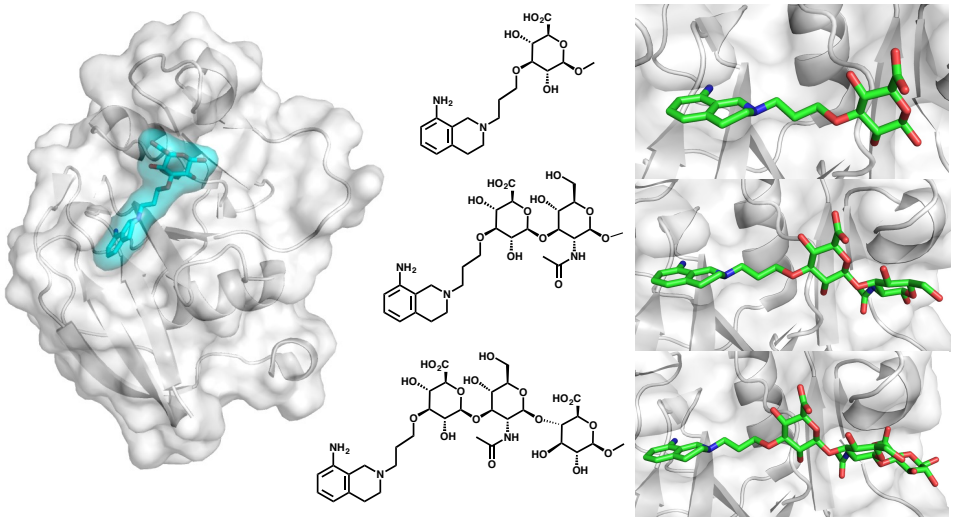


Figure 4.5 THIQ-Saccharide Conjugate Docking with CD44

*m*CD44 shown in the white cartoon and surface with the bound THIQ-monosaccharide conjugate in blue surrounded by a blue surface. The three structures (THIQ-monosaccharide, THIQ-disaccharide, and THIQ-trisaccharide conjugates) shown with their chemical structures and their docked poses in green sticks docked with *m*CD44.

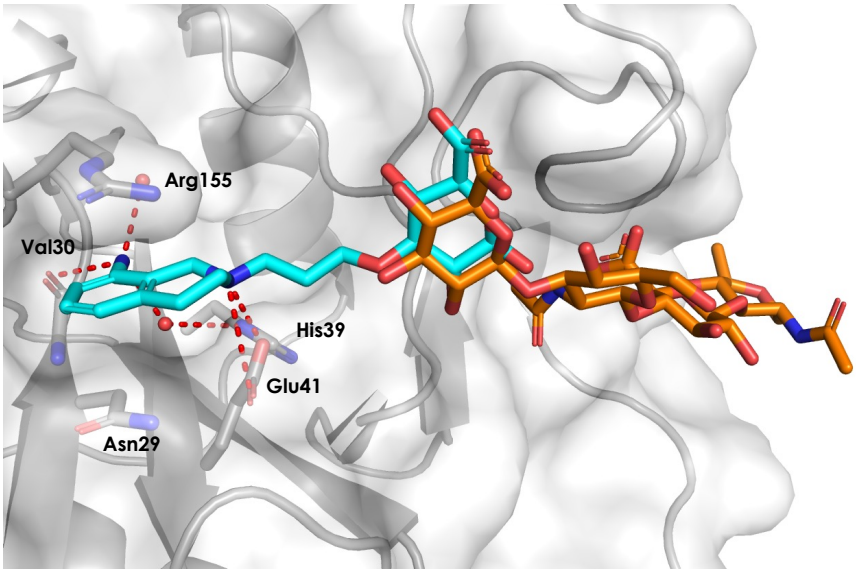


Figure 4.6 Docked THIQ-Monosaccharide Maps Onto HA-Tetrasaccharide

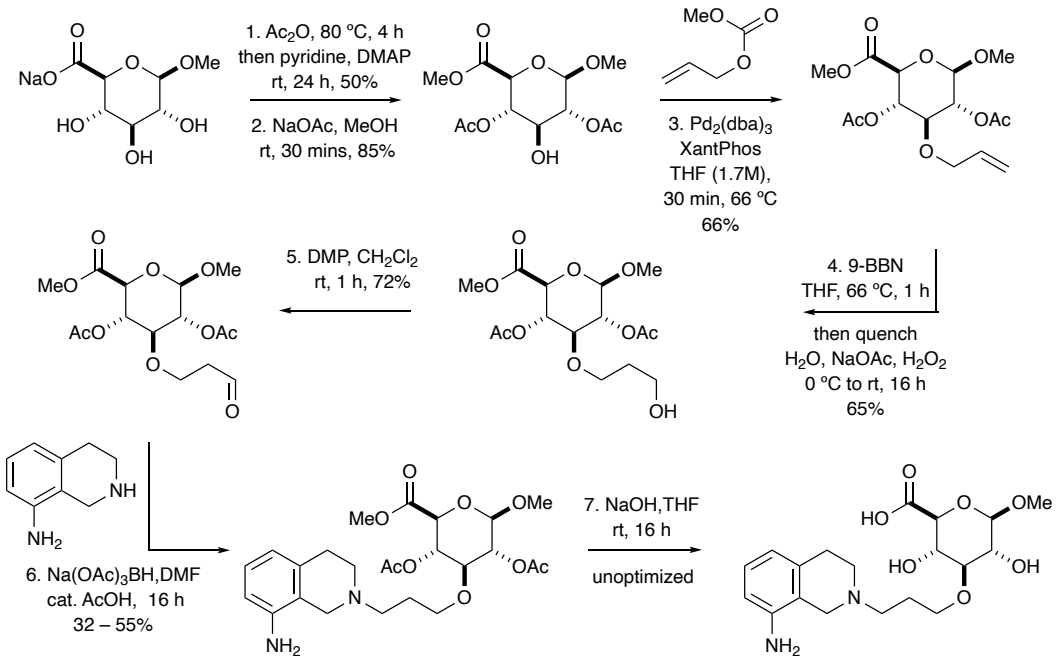
THIQ-Monosaccharide (cyan) docked into *m*CD44 (PDBid: 5bzr) places the saccharide portion onto Glc-5 of published tetrasaccharide strurcture in orange (PDBid: 4mrd).

4.3.3 THIQ-Monosaccharide Shows No Inhibitory Potential in the Immobilized-HA SPR Assay and Could Not Be Soaked into *Apo m*CD44 Crystals

With the knowledge that the THIQ-saccharide conjugates should have increased binding affinity over the unmodified saccharides, the THIQ-Monosaccharide was synthesized by Dr. Soma Maitra in Dr. Gunda Georg's lab. Synthesis involves a 7-step process with a reductive amination to link the 8-amino-THIQ to a protected Glucuronic acid and was incredibly challenging and took an extensive time to complete (Scheme 4.1). The THIQ-Monosaccharide conjugate was tested using the immobilized-HA SPR assay to determine the inhibitory potential and crystallization attempted to determine the binding modality and key interactions between the molecule and receptor. The immobilized-HA SPR was selected of the two established SPR assays in the Finzel laboratory as the goal of the project was able to selectively block HA binding to CD44 and thus causes inhibition of downstream signaling of CD44 in cancer metastasis and progression. The assay, which is competition-based (small-molecules that bind to the HABD of CD44 block CD44 from binding to the HMW-HA chip surface), was confirmed with the positive control, HA-10 (an oligosaccharide of HA containing 5 repeating units of monomeric HA). The IC_{50} of HA-10 was calculated to be 32 μM which aligned with the previously published value of 48 μM. With the functionality of the assay confirmed, THIQ-Monosaccharide was tested and found to not inhibit binding of CD44 to the chip surface as the RU level for each sensorgram, regardless of concentration of THIQ-Monosaccharide, did not change (Figure 4.7). Although an IC_{50} value could not be determined, an attempt was made to co-crystallize *m*CD44 with the THIQ-Monosaccharide to confirm the integrity of *m*CD44 and attempt to learn about the binding modality of the THIQ-Monosaccharide conjugate. The 1.84 Ångstrom resolution structure shows the protein adopts the correct tertiary structure (Figure 4.8A). *Apo m*CD44 crystals were soaked in a solution containing 10 mM THIQ-Monosaccharide for varying length of time, but even at the longest soaking duration (1 hour) the presence of the small molecule is not detected in the electron density (Figure 4.8B).

While not definitively conclusive, the inability to quantify an IC_{50} or generate a co-crystal complex leads us to reason that the THIQ-Monosaccharide conjugate

does not bind to CD44.



Scheme 4.1 Synthesis of THIQ-Monosaccharide Conjugate

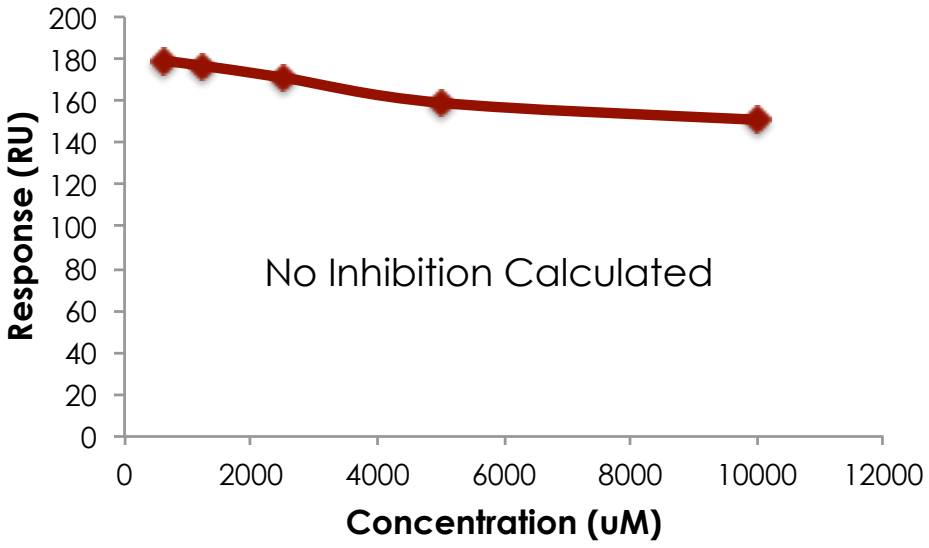


Figure 4.7 Scatterplot of Data Extracted from Sensorgrams of CD44 Incubated with THIQ-Monosaccharide Conjugate

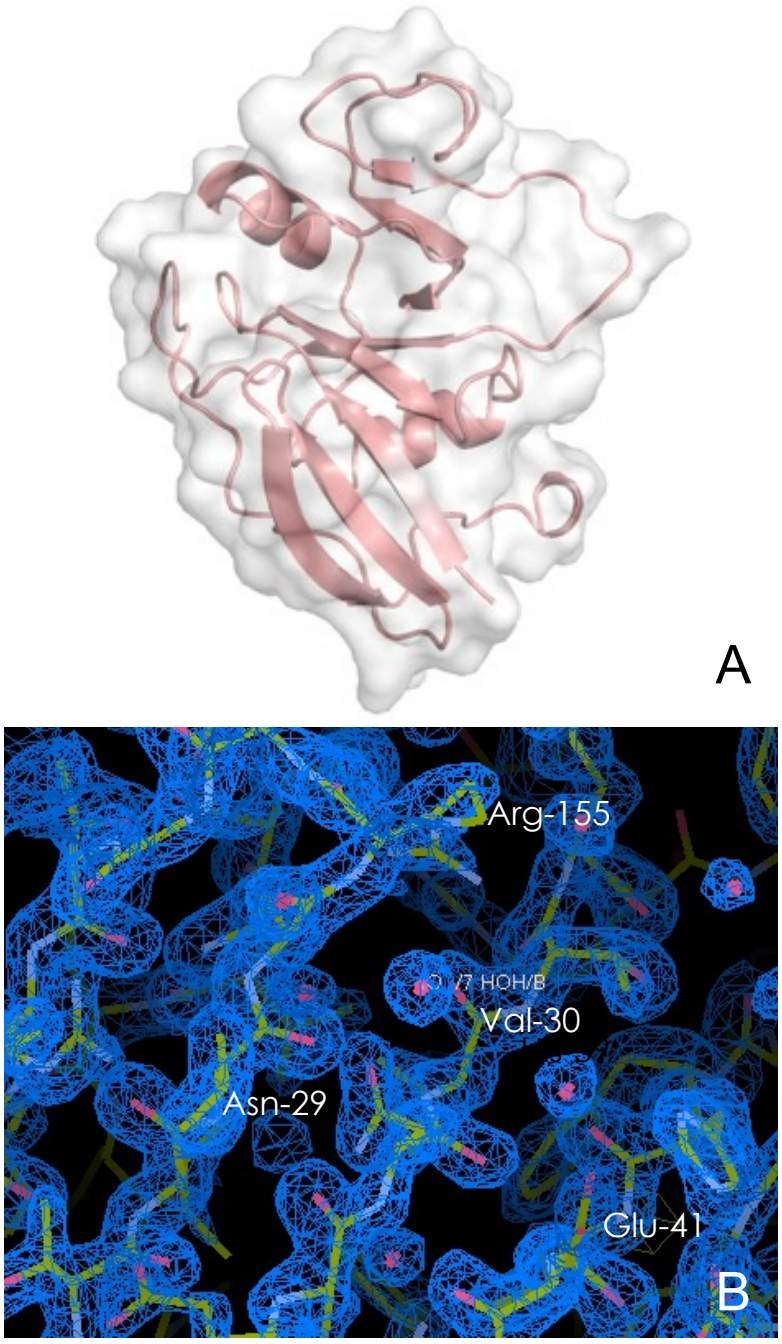


Figure 4.8 Attempted Co-Crystallization of THIQ-Monosaccharide with *m*CD44

A) Tertiary structure of *m*CD44 shown in salmon with the protein surface shown in white at 1.84 Ångstrom resolution. B) Electron density shown in blue shown of *m*CD44 showing soaking techniques to generate a co-crystal complex proved unsuccessful.

4.3.4 6 kcal/mol Energy Difference Observed in Conformation Search *in vacuo* Between Ideal Binding Conformation and Lowest Energy Conformation

Following the negative results in the immobilized-HA SPR assay and the soaking approaches to generate a co-crystal complex of *m*CD44 with the THIQ-Monosaccharide conjugate, computational efforts were once again employed to sample possible conformations of the small molecule to explain the lack of affinity observed. The THIQ-Monosaccharide conjugate was designed based on the results of the THIQ-tetrahydropyran which exhibited weak affinity but was able to bridge the small molecule binding site with the hyaluronan binding groove.[227] The three-carbon linker was selected as it gave the correct distance between the two pharmacophores (THIQ and Glc-5) and allowed the flexibility for proper binding of each component. A plausible explanation for the lack of binding affinity is that the lowest energy conformation of the THIQ-Monosaccharide was energetically further away from the ideal binding pose and that the free energy of binding was lower than this barrier. To sample the conformational poses in space, molecular mechanics or molecular dynamics can be used.[244,245] Molecular dynamics accounts for the molecular behavior in a solvated space as a function of time, but it requires a large computational effort and therefore molecular mechanics was selected.[244,245] A simple molecular mechanics calculation is doing a conformational search *in vacuo* which finds all energetically favored conformations by finding local minima on the energy curve by sampling all torsion angles of a given molecule.[241] The results of the conformational search found that the difference between the lowest energy conformation and the ideal binding conformation was approximately 6 kcal/mol (Figure 4.9). While these results do not properly simulate how the compound would behave in solution or in the presence of the protein, it points to the idea that potentially the binding energy is lower than the amount of energy required for the molecule to adopt the ideal binding pose and therefore binding would be disfavored.

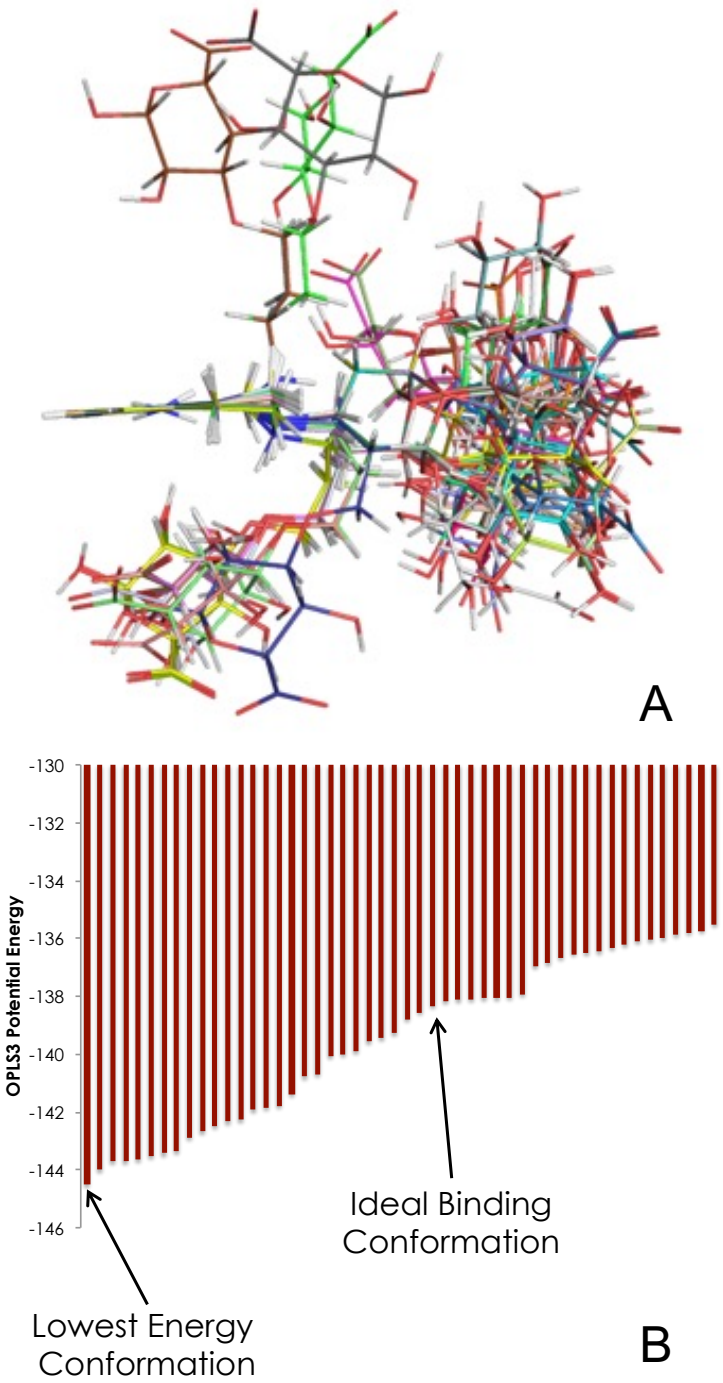


Figure 4.9 Conformational Search of THIQ-Monosaccharide Conjugates

A) 50 conformations generated by the conformational search with the THIQ moiety superimposed. B) Bar graph of OPLS3 Potential Energy vs the different conformations. The lowest energy conformation and ideal binding conformation are shown.

4.3.5 Simplification of Glucuronic Acid Allows NaG-6 to Dock Closer to Proposed Binding Site

Previous docking and scoring efforts had afforded that binding affinity should increase with increasing length of the saccharide conjugates. Since binding affinity and a co-crystal complex were not observed for the THIQ-Monosaccharide conjugate, computational efforts were employed on the THIQ-Disaccharide conjugate to guide inhibitor design. The NaG-6 binding site of the HA binding groove makes increased hydrogen bonds with NaG-6 of HA-oligosaccharides as exhibited in PDBids: 4mrd and 2jcq. The THIQ-Disaccharide conjugate and 8 mimetics of the conjugate based on deletion of functional groups around Glc-5 were examined to determine the effect of these deletions on steric hindrance/flexibility of the conjugates to allow for optimal binding position of NaG-6 (Figure 4.10). The molecules containing a deletion have the prefix "-d" followed by the functional group which has been removed in the given mimetic. The reason that the THIQ-Monosaccharide conjugate might not bind is that steric hindrance disfavors the postulated binding pose and therefore simplification should increase affinity. Initially, molecular mechanics were employed again in the form of a conformational search of the conjugate and associated mimetics, but the approach was abandoned as there were large difference in potential energy associated with nearly identical conformations. Therefore, docking and scoring using Glide was used with increased binding flexibility to better sample binding conformations. The same Xtra Precision Flexible Docking was employed on the same structure (PDBid: 5bzr) with hydrogen bonding constraints on Val30 and Glu41 with no other positional or torsional constraints. 10 conformations for each ligand were tested and then analyzed for the ability to properly dock the NaG-6 in the correct spot by rms calculations. A visual representation of the workflow for the process can be seen in Figure 4.11. Compounds that exhibit a large RMSD value are unlikely to bind and therefore can be removed from further consideration even though THIQ-d4-OH-Glc-NaG has the best docking score of any of the mimetics tested (Table 4.3). The majority of compounds then removed from further consideration all have the same removal of the hydroxyl group on position 4 of

glucuronic acid which necessitates this modification on any future inhibitor design to increase binding affinity and assure proper conformation. Of the remaining molecules two of are particular interest. The unmodified THIQ-Glc-NaG compound exhibits the best docking score which should hopefully translate into greater binding affinity. The other molecule of interest is THIQ-d2-OH-Glc-NaG as it has the lowest observed deviation from the ideal binding pose. Removal of the 2' prime hydroxyl could increase flexibility and allow for the correct positioning of the NaG-6 moiety.

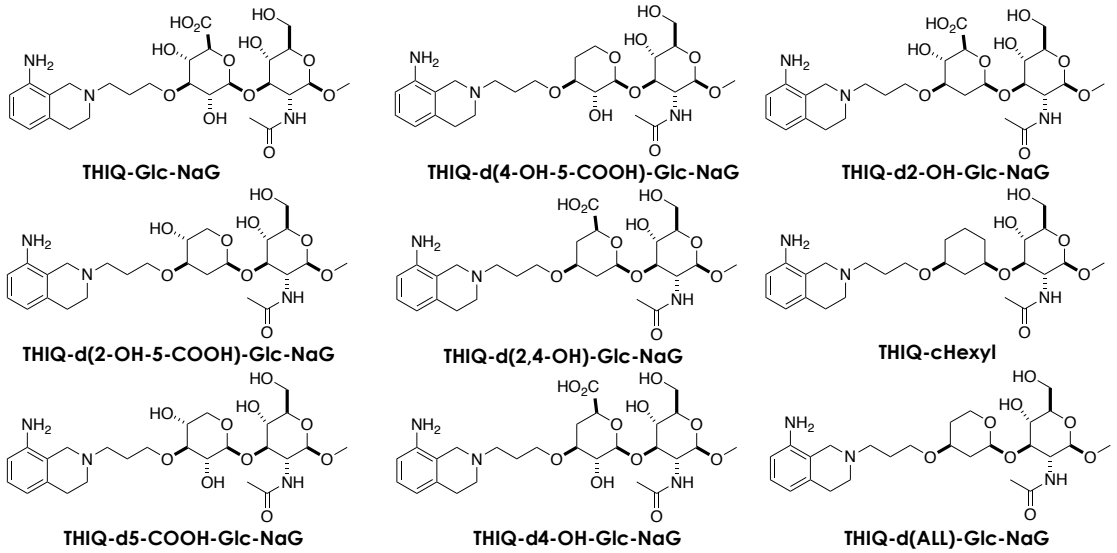


Figure 4.10 Structure and Nomenclature of THIQ-Disaccharide Mimetics

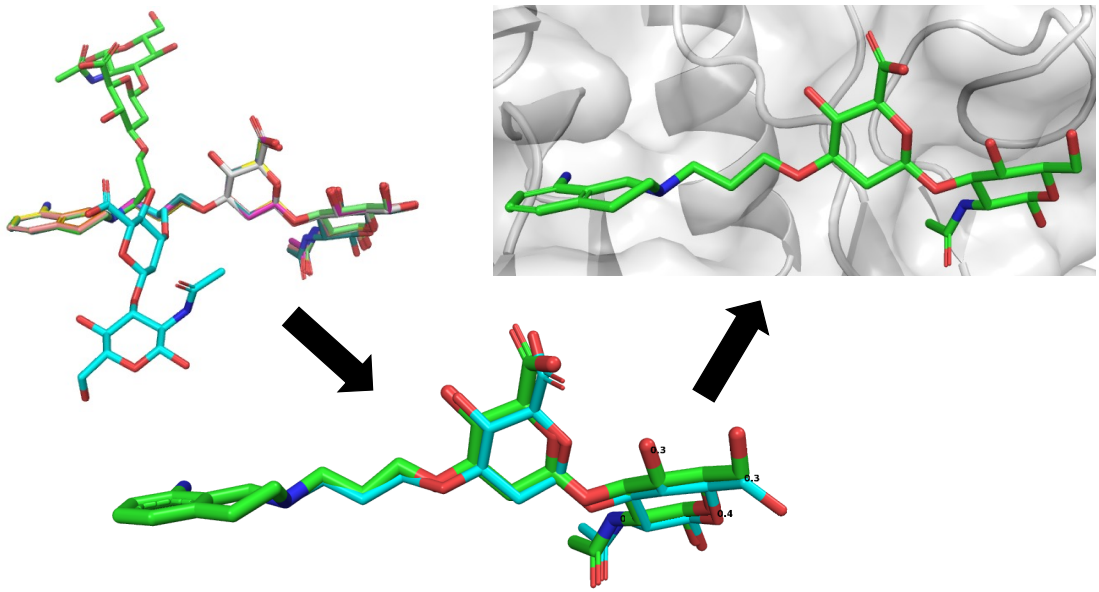


Figure 4.11 Workflow of Docking and Scoring of THIQ-Disaccharide Conjugate Mimetics

Ligands of interest were docked and scored using *Glide* with 10 conformations of each ligand. Ligands were compared to the ideal binding pose by using RMSD calculations based on the location of the NaG-6 portion of the conjugates.

Table 4.3 Comparison of Docking Scores and RMSD Values of Preferred Binding Poses of THIQ-Disaccharide Conjugate Mimetics

Compound	RMSD	Docking Score
THIQ-Glc-NaG	0.811	-7.725
THIQ-cHexyl	5.192	-7.096
THIQ-d(ALL)-Glc-NaG	0.695	-6.217
THIQ-d(4-OH-5-COOH)-Glc-NaG	1.825	-5.911
THIQ-d(2-OH-5-COOH)-Glc-NaG	0.638	-6.046
THIQ-d(2,4-OH)-Glc-NaG	0.620	-6.165
THIQ-d5-COOH-Glc-NaG	0.755	-7.088
THIQ-d4-OH-Glc-NaG	4.741	-8.033
THIQ-d2-OH-Glc-NaG	0.354	-6.842

4.4 Conclusion

Computational efforts were employed to guide synthesis and design of THIQ-saccharide conjugates. Initial efforts in docking of the saccharide units found that increasing saccharide length increases the relative docking score which follows the trend of the HA-oligosaccharides previously tested.[88] The largest change in docking scores was observed when moving from the monosaccharide to the disaccharide which is supported by previous crystallographic efforts and the hydrogen bonding network observed in the NaG-6 binding site of the HA-binding groove. The THIQ-saccharide conjugates were subsequently docked and scored with increased constraints that could be described as scoring in place instead of traditional docking and scoring. The THIQ-saccharide conjugates had better docking scores than the unconjugated saccharides with the same trend of increasing saccharide length of the conjugates increasing the relative docking score being observed. Docking scores suggest that synthesis of the disaccharide conjugate and trisaccharide conjugate should yield higher potency than the monosaccharide conjugate. The THIQ-Monosaccharide was found to exhibit no inhibition of CD44-HA interaction in the immobilized-HA SPR assay, and also a co-crystal structure of THIQ-Monosaccharide conjugate with *m*CD44 via soaking was unsuccessful. The results taken in parallel mean that the THIQ-Monosaccharide makes weak to no interaction with CD44. To explain the absence of binding, a conformational search *in vacuo* of the THIQ-Monosaccharide was performed which found a difference of 6 kcal/mol in potential energy between the ideal conformation and the lowest energy conformation. Flexible docking on the THIQ-Disaccharide conjugate found that the 4’ prime hydroxyl is critical to placement of NaG-6 and that deletion of the 2’ prime hydroxyl affords extra flexibility to adopt the ideal conformation. Future work should include synthesis of the THIQ-Disaccharide conjugate and assessment of the binding affinity and binding modality using approaches described above.

5. Biophysical Approaches to Assess Binding Affinity and Binding Modality of Verbascoside and CD44

Portion of this chapter have been excerpted from published works.

Scheme 5.1 was used with permission from: Hartmann, M.; Parra, L. M.; Ruschel, A.; Linder, C.; Morrison, H.; Herrlich, A.; Herrlich, P. Inside-out Regulation of Ectodomain Cleavage of Cluster-of-Differentiation-44 (CD44) and of Neuregulin-1 Requires Substrate Dimerization. *JBC* **2015**, *290* (28), 17041-17054.

A detailed background on the HABD of CD44 including the crystal structure bound to an HA-oligosaccharide highlighting key interactions with the groove are discussed in Chapter 1.

5.1 Introduction

Glioblastoma (GBM) is the most lethal primary brain tumor with a 3-5% 5-year survival rate.[246] General treatment for GBM includes surgery to remove majority of the tumor followed by radiation and/or chemotherapy.[247] This treatment protocol is ineffective in the majority of patients as 98% of patients will suffer from GBM reoccurrence and tolerance to radiation and chemotherapy.[248] A group of stem cell-like cells called glioblastoma stem cells (GSC) are considered to be mainly responsible to GBM reoccurrence and resistance to chemotherapy and radiation treatments.[249]

CD44 is a transmembrane glycoprotein that is a marker of GSCs and is over-expressed in a variety of tumor cells that interacts with both polymeric hyaluroan (HA) and fragments of HA.[249,250] Other than HA, substrates of CD44 include hepatocyte growth factor (HGF) and osteopontin.[251,252] Osteopontin triggers an increase in cell proliferation by causing the release of the intracellular domain through γ-secretase and localizes in the cytoplasm and nucleus with other transcription factors.[252,253] Interaction with osteopontin and HA is dependent on CD44 dimerization through disulfide bonds formation.[254] Inhibition of CD44 dimerization could therefore be an effect tumor therapeutic strategy.

Verbascoside or acetoside, as it also known, is a phenylethanoid glycoside that was first isolated from mullein but is also found in several other plant species among which is *Plantago pysillium* (Figure 5.1).[255,256] Verbascoside has been widely used in traditional Chinese medicine practices as it has various pharmacological activities including antioxidant, antimicrobial, anti-inflammatory, neuroprotective, and wound healing effects along with antiproliferative effects in cancer.[255,257] A recent study by Wang and colleagues has found that verbascoside can block CD44 dimerization and therefore act as an anti-glioma antigen through inhibition of the CD44 signaling pathway.[104] Monomers and dimers of CD44 exist on the cell surface, and dimers are stabilized by cysteine bonds within the intracellular domain (ICD) and the transmembrane domain of CD44 (Scheme 5.1).[254] Ser291 can then be dephosphorylated in the dimer complex allowing for binding of phosphorylated ERM proteins that displace merlin causing a restrictive conformation of the CD44 ICD which can be cleaved by γ-secretase.[84,253] The intracellular domain is then involved in a variety of downstream signaling pathways including the PI3K/AKT pathway contributing to cancer cell invasion and proliferation and the MAPK signaling pathway leading to cancer cell division and proliferation.[254,258,259]

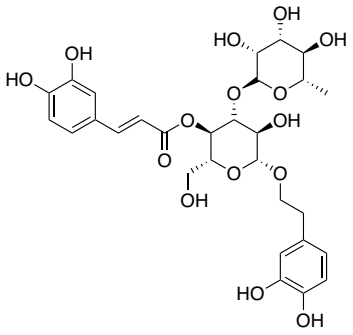
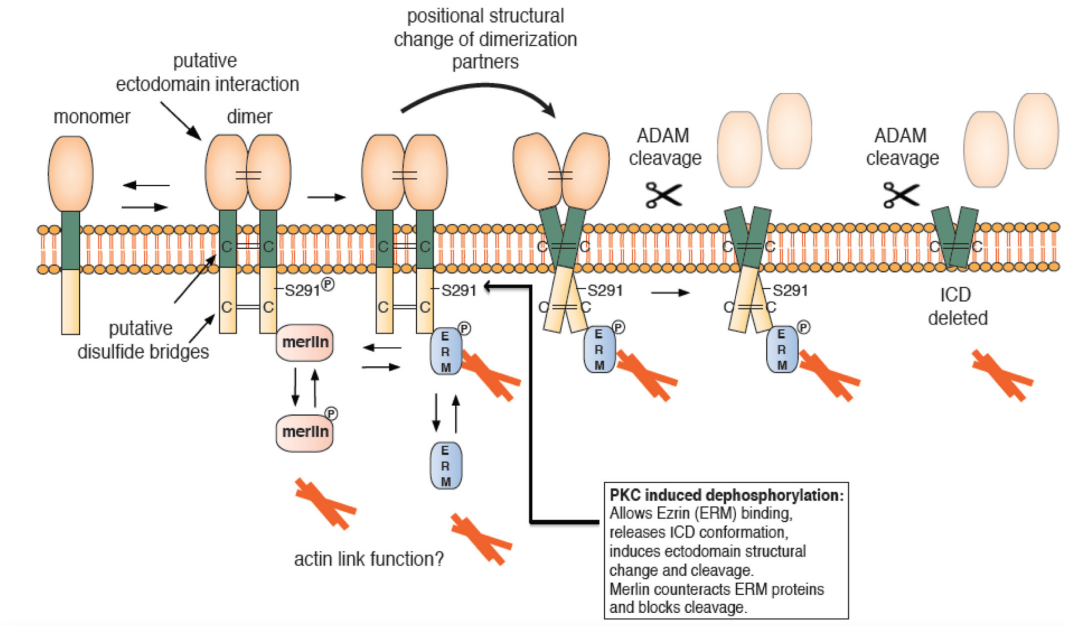


Figure 5.1 Verbascoside Molecular Structure

Using *systemsDock* and a mouse CD44 (*m*CD44) structure of the hyaluronon binding domain (2JQC) in computational modeling, verbascoside was predicted to have a higher potential binding affinity than the native ligand, HA.[85,104,260] Models also predict that verbascoside can make molecular interactions through HABD residues Ala103, Tyr109, Leu111, and Val112. These amino acids are highly conserved between *h*CD44 and *m*CD44.[85,104] Using an MST *in vitro* assay, verbascoside was found to have a lower half-maximal effective concentration (EC_{50}) than verbascoside and osteopontin.[104] HEK293 cells-transfected with CD44-YNE/CD44-YCE were used in a biomolecular fluorescence complementation (BiFC) assay to quantify the effect of

verbascoside on CD44 dimerization.[104] Verbascoside was found to decrease dimerization in a dose-dependent fashion.[104] U251MG cells treated with 60 μM verbascoside had lower levels of the cleaved CD44 ICD thus showing inhibition of the CD44 signaling cascade.[104] Finally using a xenograft mouse model, verbascoside was found to cause a reduction in tumor growth and prolonged survival suggesting that verbascoside may be an effective treatment of mouse GBM.[104]



Scheme 5.1 CD44 Cleavage Regulation by Dimerization

Our work with verbascoside sought to further understand the molecular interactions of CD44 with verbascoside to guide future inhibitor design aimed at increasing selectivity and potency. Verbascoside is postulated to interact with the hyaluronan binding groove of CD44 HABD based upon computational methods, but no structural or biophysical data supported this binding modality. The Finzel laboratory has successfully characterized the binding affinity, inhibition and structure of a series of small molecules and HA-oligosaccharides and these tools can be employed to confirm that verbascoside actually interacts with the CD44 HABD. We applied three biophysical assays to gain an understanding of the molecular interactions. These findings were intended to be used along with crystallographic data of the co-crystallized complex to support or refute the binding modality observed in Wang and colleagues computational docking studies. Verbascoside was

found to cause a small destabilization of CD44 HABD using differential scanning fluorimetry, but neither of the surface plasmon resonance assays were able to provide evidence of a direct interaction between verbascoside and the HABD of CD44.

5.2 Experimental

5.2.1 Reagents

Full-length human CD44 (*h*CD44) cDNA was a gift of Dr. James B. McCarthy at the University of Minnesota. Rosetta 2(DE3)pLysS cells were purchased from EMD Biosciences (Gibbstown, NJ) and competent cells were made using a Mix and Go! E. coli Transformation Kit from Zymo Research (Irvine, CA). HiPrep 16/60 Sephacryl S-100 HR was obtained from GE Healthcare (Piscataway, NJ). High molecular weight hyaluronan (HMW-HA) sodium salt from rooster comb was acquired from Sigma-Aldrich (St. Louis, MO). CM5 and SA Biacore Sensor Chips were purchased from GE Healthcare (Piscataway, NJ). L-Arginine, reduced/oxidized glutathione, coupling reagents, and other general buffer reagents were from Sigma-Aldrich (St. Louis, MO). Hydrazide biotin linker for making biotinylated HA and PCR Plates (96-well, low profile, non-skirted, white) for DSF experiments were from Thermo Scientific (Rockford, IL).

5.2.2 Expression and Purification of *h*CD44

*h*CD44 and *m*CD44 HABD was cloned, expressed, and purified as previously described in Chapter 4.[88]

5.2.3 Differential Scanning Fluorimetry

Differential scanning fluorimetry experiments were conducted on a BioRad CFX96 Well System (Hercules, CA) using 96-well PCR plates and previously published protocols.[208]. To determine the optimal protein concentration for use in the DSF experiment, a scan was completed on a two-fold dilution series of *h*CD44 HABD (0.5-0.01 mg/mL). To the wells was added protein at varying concentration, Sypro-Orange (1:500 dilution of 5000x), and DSF Buffer (20 mM Tris pH 8.0 and 150 mM NaCl) to a final volume of 40µL. The DSF scan was completed using 0.5°C

steps (25-80°C) while holding for 30 seconds at each temperature increment. The resulting fluorescence curves were fitted using a Boltzmann model for non-linear regression in the curve fitting software: Bio-Rad CFX 2.0.[209,210] A concentration of 0.1 mg/mL was used for subsequent experiments using the least amount of protein while still getting a robust fluorescence signal. To determine the effect of verbascoside on protein stability, several concentrations of verbascoside (5-0.625 mM), solubilized in the SEC buffer, were added to the well components described above and tested in triplicate. Effect on protein stability was calculated by taking the denaturation temperature of *h*CD44 in the presence of verbascoside and subtracting that by the denaturation temperature of *apo h*CD44.

5.2.4 Immobilized CD44 HABD SPR Assay

Determination of verbascoside binding affinity for *h*CD44 HABD was performed on an immobilized *h*CD44 HABD surface using Surface Plasmon Resonance (SPR) on a Biacore S200 (GE Healthcare) as previously described.[88] A CM5 Biacore Sensor Chip was activated with 0.4 M 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride and 0.1 M *N*-hydroxysuccinimide prior to flowing *h*CD44, that had been dialyzed against 10 mM MES pH 4.4 overnight, until a 3500 Response Units (RU) level was achieved. The remaining reactive carboxymethyl groups were blocked with 1 M ethanolamine. Verbascoside in a PBS buffer containing 0.005% Surfactant P-20 was passed over the chip via a two-fold dilution series (10 mM – 78 μM) at 30 μL/min for 60 seconds. Bound verbascoside was dissociated from the chip surface using the same running buffer over 60 seconds at 30 μL/min to return the chip back to initial baseline. RU values were plotted against log[concentration of verbascoside] and fitted with a non-linear regression using GraphPad Prism.[218]

5.2.5 Immobilized High molecular weight hyaluronan (HMW-HA) SPR Assay

A competition-based HA-binding assay was performed to determine the inhibitory potential of verbascoside using SPR on a Biacore S200 (GE Healthcare). HMW-HA sodium salt from rooster comb was biotinylated using hydrazide biotin and immobilized onto a SA Biacore Sensor chip using previously published methods.[88,242] In brief, biotinylated HMW-HA was flowed over the streptavidin chip surface until the response signal stabilized at 115 RU as monitored with the Biacore S200. For the competition-based assay, a fixed concentration of *h*CD44 HABD (~17 μM) was incubated for 10 minutes on a plate shaker at room temperature with a two-fold dilution series of verbascoside (10 mM – 78 μM) in HEPES-EP buffer. CD44-verbascoside mixtures were passed over the immobilized HMW-HA chip surface at 30 μL/min for 60 seconds followed by 2 minutes of a dissociation phase using HEPES-EP buffer. Levels of CD44 bound to the chip surface were determined by taking the RU level at 5 seconds before the end of the association phase.

5.3 Results and Discussion

5.3.1 Verbascoside Causes Small Shift in *h*CD44 Denaturation Temperature

Differential scanning fluorimetry (DSF) experiments were conducted as a fast and inexpensive experiment to assess the effect of the small-molecule on protein folding and stability. The T_m, melting temperature, is the point where the amount of native folded protein is equal to that of the denatured or unfolded protein. A T_m for *h*CD44 in the absence of a small-molecule was calculated to be 45.7°C which is slightly higher than previously published results (45°C), but it confirms that the protein is properly folded and behaving similarly.[227] The T_m is not as important as changes in the T_m upon ligand binding that correlate with the effect of binding on stability of the protein of interest. The highest concentration of verbascoside caused a 1.7°C decrease in the unfolding temperature of the protein (Figure 5.2A). In drug discovery and fragment screening using DSF, generally, positive shifting molecules are prioritized.[214] A negative shifting ligand may have

a destabilizing effect on the tertiary structure of the protein, or stabilize a non-native state of the protein. Several researchers have published negative shifting ligands/fragments that have preceded successes in their medicinal chemistry campaigns.[261–263] Verbascoside was tested at five different concentrations in a two-fold dose response and the change in thermal stability was found to be proportional to the concentration of verbascoside present (Figure 5.2B). This same trend follows previous observations of stabilizing compounds.[264–266] 1.7°C is a relatively small shift, but verbascoside causes a destabilization of the *h*CD44 HABD.

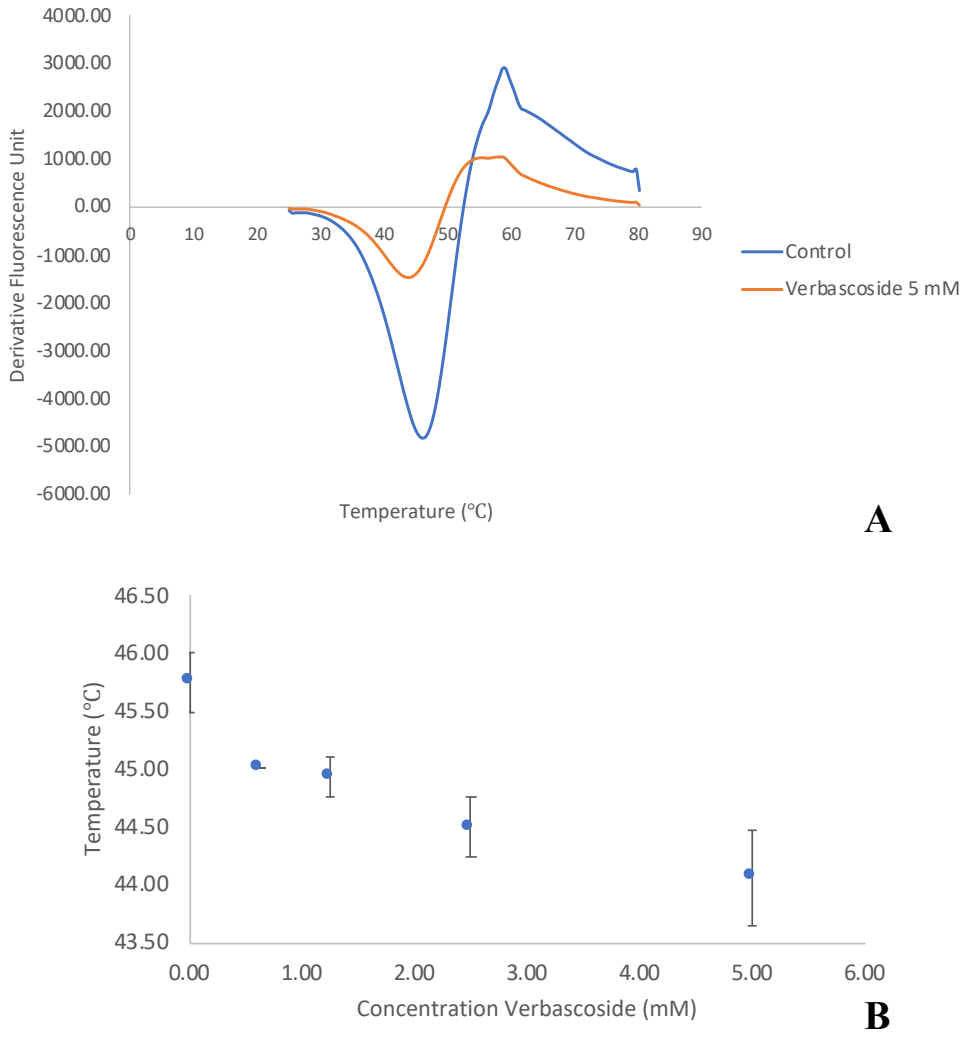


Figure 5.2 Verbascoside Effect on *h*CD44 Protein Stability

A) First-derivative plot of fluorescence versus temperature with *apo* protein shown in blue and protein + verbascoside (5 mM) in red. B) Dose response on protein stability versus varying concentration of verbascoside. Standard deviation for the melting temperatures are shown with the black error bars for a dataset of N=8.

5.3.2 Verbascoside Binds Weakly to CD44 HABD

To determine the binding affinity, K_D, of verbascoside for *h*CD44 HABD, a previously published immobilized HABD SPR assay was performed.[88] To confirm the integrity of the chip surface, an HA-oligosaccharide, HA-10 (5 repeating monomeric units of the HA disaccharide), was

tested and the resulting K_D value aligned with previously published values. Verbascoside in a two-fold dilution series was flowed over the chip surface and the small-molecule caused a change in the refractive index which can be quantified as an increase in the response units. Increasing concentration of verbascoside causes a higher response denoting that more small-molecule is bound to the chip surface (Figure 5.3). For determination of the K_D, RU values taken 5 seconds prior to the end of the association phase were used and plotted in GraphPad Prism. The software gives a projected K_D of 24.9 mM (the term projected is applied as the value is higher than the highest concentration tested in the experiment). The published solubility in water or DMSO of verbascoside is 100 mg/mL (~160 mM) and to reach saturation in the SPR assay 10x the K_D needs to be tested and this concentration (~250 mM) is not possible to reach in solution. Increasing the concentrations tested in the assay could help get a better estimate of the K_D, but the binding affinity is still low.

An odd phenomenon observed in these spectrographs is the high RU signals which is equal to the amount of bound analyte in each experiment. In the assay as the concentration of verbascoside is increased, the RU signal also increases and does not reach a saturation level. The most logical explanation for this occurrence is that verbascoside binds with a stoichiometry greater than 1:1. The maximum RU is given in Equation 5.1 and is dependent on the amount of ligand (*h*CD44 bound to the chip surface), the molecular weight of the ligand and the analyte (verbascoside), and the stoichiometry of binding.[267] An initial stoichiometry of binding of 1:1 was used in calculations as that is the ratio that was assumed in the virtual screen.[104] Entering implicit values into Equation 5.1 (R_{max} is the maximal RU signal, R_{ligand} is the amount of immobilized ligand, M_r is the molecular weight, and V_{ligand} is stoichiometry of binding), the maximal RU for the experiment should be approximately equal to 124 RU. The stoichiometry of binding, according to the 10 mM sensorgram of verbascoside (RU = 660), is at least 1:5 (protein:analyte) which means there are multiple spots on the protein that verbascoside is able to interact with more than one site on CD44 HABD. Crystallographic efforts with *m*CD44 have shown that the HABD groove can accommodate up to an 8 saccharide units.[85] Verbascoside or caffeoyl phenylethanoid glycoside

containing a disaccharide of β-ethyl-O-α-L-rhamnopyranosyl-β-D-glucopyranoside would mimic a dually modified HA-dimer and therefore it is possible that the groove could accommodate more than one molecule of verbascoside. In the assay, verbascoside interacts with *h*CD44 but orthogonal assays and structural biology efforts are needed to understand the binding modality and the number of binding sites on *h*CD44 HABD that can interact with verbascoside.

Equation 5.1 Maximum Theoretical RU for a Given Ligand and Analyte

$$R_{max} = R_{ligand} * \frac{Mr_{ligand}}{Mr_{analyte}} * V_{ligand}$$

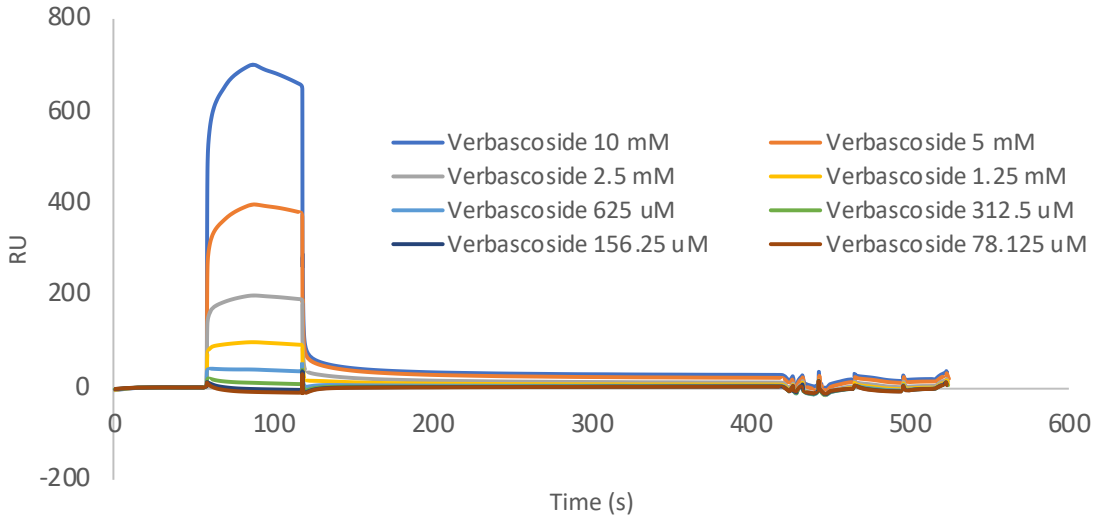


Figure 5.3 Sensorgram of Verbascoside in Immobilized HABD SPR Assay

5.3.3 Verbascoside Does Not Decrease Bound *h*CD44 HABD in a Competition SPR Assay

Since an accurate determination of the K_D was not obtained, an orthogonal approach to quantify the binding interaction of verbascoside with *h*CD44, a previously established immobilized-HA SPR competition assay, was employed.[88] Various protein concentrations were flowed over the chip surface to establish a dose-dependent effect on the RU signal observed that reflects CD44 HABD binding to the chip surface and confirms the immobilization of HMW-HA on the chip surface. The premise of the assay is that small molecules that bind to the HABD of

*h*CD44 will inhibit binding of its natural substrate, HMW-HA (the active chip surface), so a decrease in binding of *h*CD44 to the surface will be observed relative to *apo* protein. To test the competitive inhibition, an HA-oligosaccharide, HA-10, was used as a positive control. HA-10 was incubated with the protein prior to being flowed over the chip surface and inhibits binding of CD44 to the chip surface at increasing concentration of HA-10. An IC50 value, which can be calculated from the resulting sensorgram data, of 32 μM was obtained, which aligns well with previously published data.

With the assay working, verbascoside was tested in a two-fold dose-response assay. The sensorgrams shows that verbascoside at the highest concentrations tested appear to decrease the amount of CD44 bound to the chip surface (Figure 5.4A). The highest concentration of verbascoside (10 mM) has a negative sensorgram; the RU level is less than the initial baseline signal, which means that the analyte is interacting more strongly or non-specifically with the reference chip surface containing unmodified streptavidin than it is with the active chip surface. SPR experiments detect binding levels for an analyte and ligand of interest relative to the analyte only binding levels. SPR chips consists of lanes and an experiment consists of injecting the same sample over two lanes: one the active chip surface with an immobilized ligand (Biotinylated HMW-HA in this experiment) and the other the reference chip surface which does not have immobilized ligand (streptavidin only). The sensorgrams in Figure 5.4A are the result of taking the sensorgram corresponding to the analyte effect on the active chip surface and subtracting the sensorgram of the analyte over the reference chip surface. The subtraction step should, for the most extent, remove any effects of buffer mismatch (analyte conditions do not align with running buffer), fluidic effects, and non-specific binding that occurs in both the reference and active surface, and therefore, allow for analysis of the isolated effect of the analyte interacting with the ligand. Upon closer examination of the binding to the reference chip surface, a tool included within the Biacore Evaluation Software, verbascoside at increasing concentration interacts more strongly with the reference chip surface (Figure 5.4B).[268] The Biacore Software determines the reference binding levels by comparing the

RU levels of the reference chip surface relative to the lowest concentration of analyte tested (any change in buffer or flowing a small molecule will cause a change in the refractive index regardless if it interacts or not). The lowest concentration of analyte is set at zero for the reference binding level and if molecules interact more strongly with the reference chip surface this change in RU is quantified as the binding reference level. For accurate biophysical interpretation and analysis, the analyte should interact consistently with the reference chip surface regardless of concentration. Since verbascoside is interacting with the reference chip surface, an additional mathematical step was taken to attempt to understand the effect of verbascoside on the inhibition of CD44. The reference binding levels (described above and shown in Figure 5.4B) were added to the RU levels taken 5 seconds prior to the end of the association phase in the sensorgrams below (Figure 5.4A). This mathematical manipulation of the data should remove the non-specific interactions on the reference chip surface and isolate CD44 interaction with the immobilized-HA surface as caused by verbascoside concentration. The majority of the two-fold dose-response (78.125-1250 μM) is relatively flat (no decrease in bound *h*CD44) but there is a slight decrease in RU at the highest concentrations tested (Figure 5.4B). Due to the shape of the curve (follows a more linear fit than an expected sigmoidal curve) an accurate IC_{50} value could not be determined.

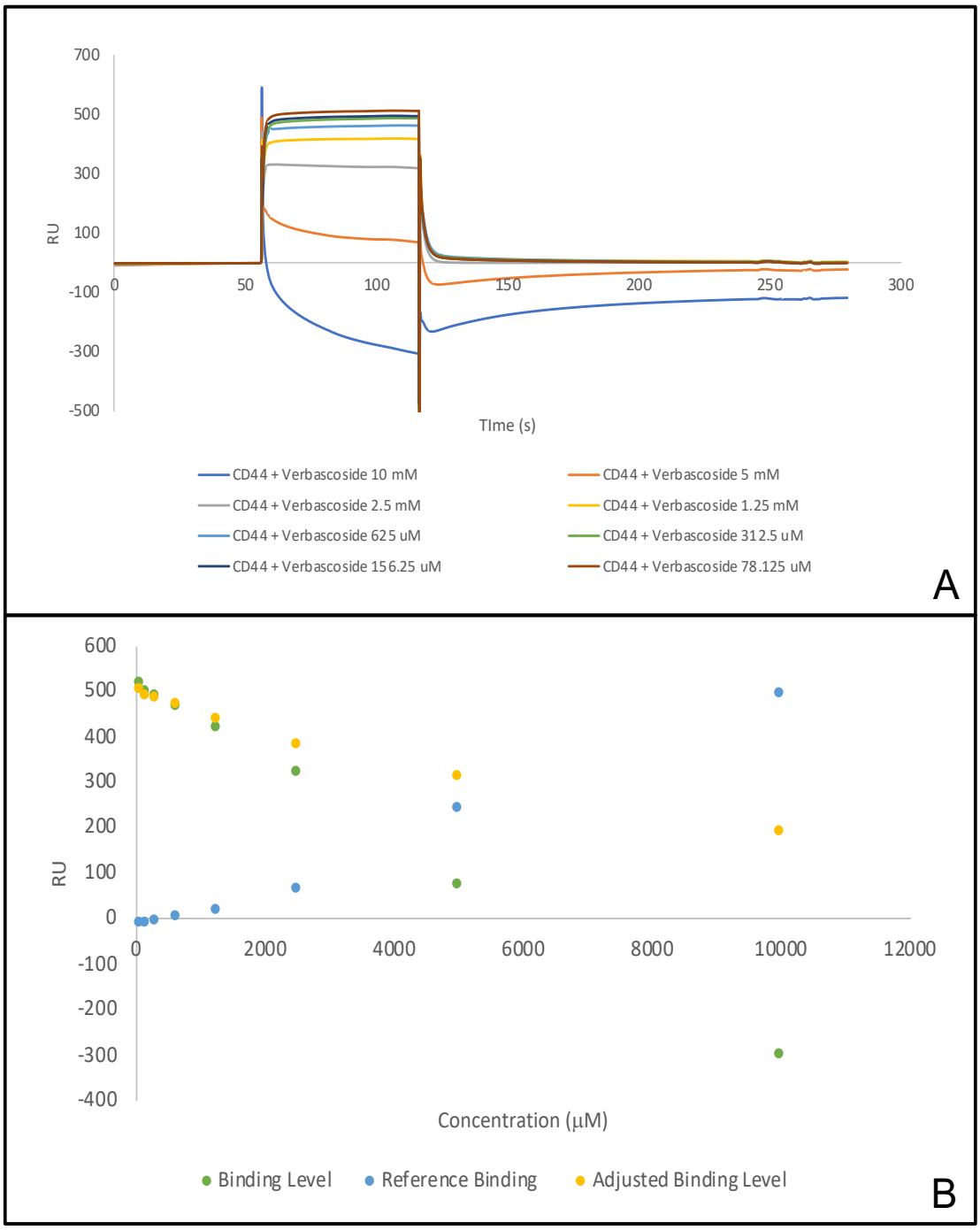


Figure 5.4 Verbascoside Immobilized-HA SPR Assay Results

A) Sensorgram of verbascoside in the immobilized-HA SPR assay with the legend shown below the graph. B) Scatterplot of binding level (green), binding to the reference chip surface (blue), and adjusted binding level (yellow).

5.4 Conclusion

Verbascoside was tested using two biophysical techniques in a total of three different assays: differential scanning fluorimetry, an immobilized-HABD surface plasmon resonance assay, and an immobilized-HA surface plasmon resonance. Verbascoside was found to nominally decrease the protein unfolding temperature of *h*CD44 relative to *apo* protein, likely by binding to and stabilizing a non-native (or partially unfolded) state of the protein. To further quantify and qualify verbascoside interaction with *h*CD44, two surface plasmon resonance assays were employed to determine the binding affinity, K_D, and inhibitory potential, IC_{50}. Verbascoside caused an increase in RU level that was proportional to the amount of analyte flowed over the chip surface. While a K_D of approximately 25 mM can be projected, an accurate K_D could not be determined as the projected K_D was greater than the highest concentration tested in the assay. Solubility issues with verbascoside make an accurate determination of the K_D not possible. The RU values observed at the highest concentration were greater than the calculated maximum RU leading to a conclusion that verbascoside and CD44 may not interact with a 1:1 stoichiometry. Further investigation is needed to support this notion. Finally, verbascoside was tested in the immobilized-HA SPR assay to determine the IC_{50} value. Verbascoside interacted in a dose-dependent manner with the reference chip surface, where no HMW-HA has been immobilized. Taking into account the non-specific interaction with the reference chip surface, it was found that only at the highest concentrations of verbascoside tested was there a slight decrease in the amount of bound *h*CD44. Through our biophysical studies we were not able to gain further insight and understanding into the interaction between verbascoside and *h*CD44. Our findings show weak CD44-verbascoside interactions in two of the assays tested which does not disprove or support the findings published by Wang and colleagues.

6. A Novel Surface Plasmon Resonance Assay and Probability Function to Assess Modified-HA Inhibitors of CD44

Author contributions:
Kunli Liu in Dr. Xuefei Huang's laboratory at Michigan State University completed the synthesis of the modified-HA molecules. The synthesis of these molecules is not discussed in detail but the resulting compounds are used in this study. The ELISA experiment and data analysis was completed by Kunli Liu and permission was given to include the data in this chapter.

An in-depth look at the extracellular matrix in cancer, the signaling molecule HA, the family of hyaldherins, the HA-binding protein CD44, and the structure of the HABD and its interaction with HA is included in Chapter 1.

6.1 Introduction

Hyaluronan (HA), repeating units of glucuronic acid (Glc) and N-acetyl glucosamine (NaG), is an integral component of the extracellular matrix (ECM) where it provides structural and functional integrity to cells and tissues.[89] High concentrations of HA are needed in many organs to provide structural framework and space filling for cells to migrate.[63] To shape the molecular environment in which they live, cells must orchestrate a balance between production and turnover of HA with roughly one-third of the total body HA concentration being replaced daily.[269–271] Polymeric-HA (1,000 to 10,000 kDa) is catabolized to fragments (10 to 100 kDa) that can be recycled into newly synthesized HA.[89,90] Certain pathological conditions such as cancer causes an increase in the turnover of HA where these fragments of HA serve as signaling molecules in the matrix[91,270,272] and have been linked to cell differentiation[72], angiogenesis[73,91], tumorigenesis[74,75], and resistance to cancer chemotherapy[269,273].

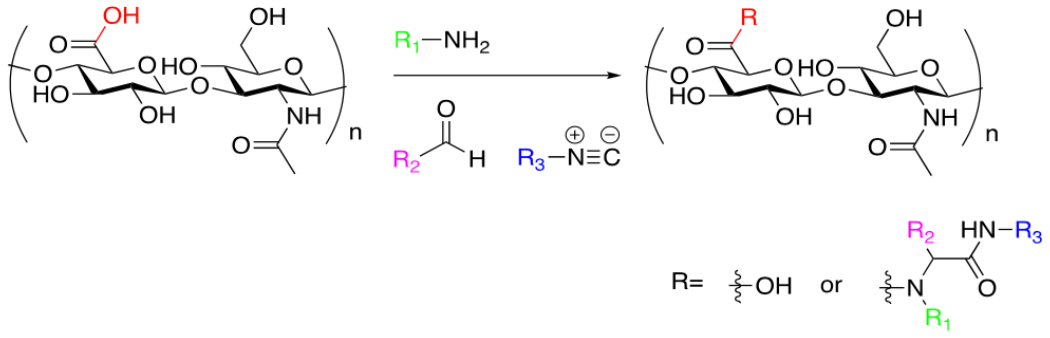
CD44 is of the family of hyaluronan-binding proteins, hyaladherins, that has been discussed extensively in the previous two chapters (Chapter 4 and Chapter 5) and in the introduction chapter (Chapter 1).[77] Previous structural data has shown that CD44 is a challenging drug target. The murine CD44 HABD (*m*CD44) complex with an HA oligosaccharide has an extended HA

binding site (surface area greater than 800 $Å^2$) with molecular binding stabilized by many weak interactions with at least seven consecutive saccharide units of HA.[85] The protein-polysaccharide complex resembles protein-protein interactions that have been a difficult target for medicinal chemists to disrupt with small-molecules. Structural biology and biophysical fragment screening in the Finzel laboratory resulted in the discovery of an inducible small-molecule subsite near the hyaluronan binding groove.[88,227] One approach to potentially target this challenging target is linking a small-molecule with known affinity for CD44 to an oligosaccharide of HA as a means of increasing binding affinity over HA and imparting selectivity for CD44 over the other proteins of the family of hyaladherins (extensively described in Chapter 4). Another approach for targeting CD44 is by modifying full-length HA to screen for conjugate molecules with larger binding affinity and inhibitory potential relative to the unmodified HA.

To target the hyaluronan binding domain of CD44, Dr. Xuefei Wang and Kunli Liu at Michigan State University designed a series of modified-HA molecules using an Ugi reaction.[274] The Ugi reaction is a four-component reaction that requires carboxylic acid-, amine-, aldehyde-, and isocyanide-containing reagents.[275] The glucuronic acid (Glc) portion of HA contains a carboxylic acid moiety that can be modified with different amine, aldehyde, and isocyanide combinations to generate a library of HA derivatives (Scheme 6.1).[274] The Ugi reaction does not modify every carboxylic acid and therefore the loading percentage, or ratio of modified to unmodified carboxylic acid moieties, can be determined by comparison of the unmodified full-length HA (FL-HA) to the modified HA using nuclear magnetic resonance (NMR).[274] Polymeric HA is the composite of many different lengths of HA and the Ugi reaction can occur with varying frequency and pattern therefore resulting in a diverse and varied population of modified HA molecules: molecules within a given sample are not the same. The library of modified HA molecules was then tested using a competitive ELISA assay with CD44 and two other hyaladherins, LYVE-1 and HARE, to determine binding affinity relative to the unmodified native substrate, HA,

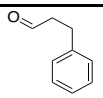
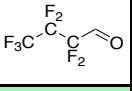
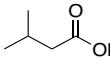
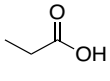
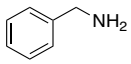
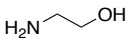
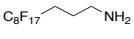
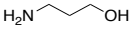
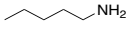
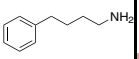
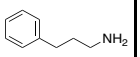
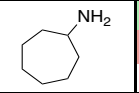
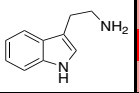
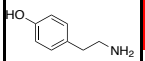
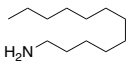
in hopes of finding a modification that has increased affinity over HA and binds tighter to CD44 than the other hyaladherins to generate a selective molecule (Table 6.1).[274] The results of the ELISA found that the aldehyde 2, formaldehyde, is probably the privileged aldehyde for the HA derivatives and amines E, F, G, I, and K (where F, G, I, and K are all aromatic containing amines) are the tightest binders of CD44.[274] The ELISA assay did not quantitative binding and therefore the team at Michigan State reached out to the Finzel laboratory to employ our published surface plasmon resonance (SPR) assays to quantify the K_D and IC_{50} values of a series of modified-HA molecules. Three modifications at different loading percentages were sent to the Finzel laboratory for biophysical characterization and comparison of the effect of loading percentage on binding affinity and inhibition potential. The three modifications that were selected for analysis were A2a, F2a, and G2a (Table 6.2). Both F2a and G2a exhibit increased and comparable binding affinity for CD44 as calculated by the ELISA assay: 38%±1%, and 38%±0%, respectively. A2a on the other hand has lower affinity for CD44 than the natural substrate, HA (-14%±6%). The two modifications with improved binding affinity were selected to confirm the binding affinity in an orthogonal assay and to assess any difference in the two modifications with respect to their effects on CD44. The A2a modification was selected as it could be used as a control compound to show a compound that binds with lower affinity for CD44. The three modifications all have the same aldehyde, carboxylic acid, and isocyanide reagents and therefore the contribution of the specific amines can be assessed. Initial concern of these molecules was that an accurate binding affinity could not be determined as the starting material, HA, is a distribution of species with a mean molecular weight and the loading percentages are variable between compounds in the library. We therefore sought to compare the relative response of HA and the modified-HA molecules at the same fixed running concentration (μg/mL) to determine the relative binding affinity and the relative inhibitory potential in two SPR assays: an immobilized HABD SPR assay and an immobilized-HA SPR assay. Due to the polymeric nature of the molecules, we encountered flow and avidity issues which made determining the binding affinity impossible. We developed a ligand reference subtracted SPR assay

which decreased the noise of the polymeric ligands thus allowing a clearer assessment of the effects of the modified-HA molecules on CD44-HA interaction. We observed a decrease in the inhibitory potential with increasing loading percentage which is contradictory to traditional small-molecule inhibition (more small-molecule equals greater inhibition) and therefore developed a probability function to explain this relative decrease. Our formula and known structural biology work suggest that CD44 HABD is only able to accommodate one modification per 8 saccharide units (the length which the HABD groove can accommodate) and with higher loading percentage we observe a higher likelihood of smaller distances between modifications.



Scheme 6.1 Ugi Reaction to Modify Glucuronic Acid of HA

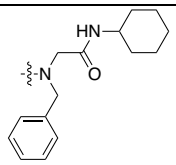
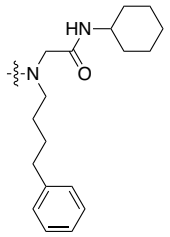
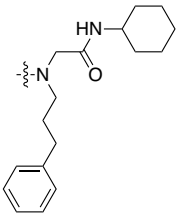
Table 6.1 ELISA Data for Modified-HA Molecules Against Three Hyaladherins

		1	2	3	4	5
						
A		40%±0% 41%±0 N.D.	-14%±6% -7%±1% 5%±5%	-13%±6% 3%±3% N.D.	-10%±0% N.D. N.D.	-13%±4% 11%±0% N.D
B		-8%±0% N.D. 5%3±2%	3%±11% N.D. N.D.	-9%±7% N.D. -9%±3%	-14%±6% 8%±0% N.D.	-11%±0% 7%±2% 3%±10%
C		N.D. N.D. N.D.	23%±14% 23%±1% 12%±7%	34%±5% 24%±3% N.D.	N.D. N.D. N.D.	N.D N.D N.D
D		-10%±2% N.D. 5%±11%	2%±0% N.D. 4%±1%	-6%±1% N.D. N.D.	-23%±3% 2%±2% -3%±8%	-18%±3% 3%±1% N.D
E		-7%±1% N.D. N.D.	11%±1% 4%±1% -3%±3%	0%±1% N.D. N.D.	-17%±0% 6%±6% -2%±4%	-15%±0% 11%±2% 0%±0%
F		1%±1% N.D. 11%±2%	38%±1% 30%±1% -5%±3%	1%±2% N.D. -4%±2%	-6%±2% N.D. N.D.	-8%±1% N.D -2%±3%
G		4%±1% N.D. 0%±0%	38%±0% 22%±0% 5%±2%	-6%±0% N.D. -4%±1%	4%±2% N.D. 21%±5%	-5%±2% N.D 0%±2%
H		-13%±5% 6%±1% N.D.	-26%±2% 8%±0% 3%±4%	-7%±1% N.D. -2%±1%	-23%±0% 12%±2% -2%±7%	-10%±0% 13%±1% -4%±3%
I		9%±1% 28%±0% 12%±4%	22%±0% 11%±1% 0%±8%	-1%±1% N.D. -10%±1%	4%±6% N.D. 15%±3%	-8%±1% N.D 14%±3%
K		30%±2% 32%±0% N.D.	23%±1% 4%±1% 7%±3%	3%±1% N.D. 9%±2%	-22%±4% 19%±4% 13%±7%	-25%±3% 5%±2% 8%±3%
L		8%±1% 35%±1% 30%±0%	36%±0% 33%±1% 24%±2%	29%±1% 27%±0% N.D.	-18%±5% N.D. N.D.	20%±1% 33%±1% N.D

Inhibition Difference	<-15%	-15% to -5%	-5% to 5%	5% to 15%	> 15%

If inhibition difference is positive, the HA derivatives have enhanced affinity toward the hyaladherin tested (red highlighting) and if it is negative, it has decreased affinity (green highlighting). N.D. represents the data is not available. Compounds were tested against: CD44, LYVE-1, and HARE (data down the column).

Table 6.2 Structure (R-Group) of Modified-HA Molecules

Molecule	Structure (R group)	Relative Inhibition Difference as Calculated by ELISA
A2a		-14%±6%
F2a		38%±1%
G2a		38%±0%

6.2 Experimental

6.2.1 Reagents

Full-length human CD44 (*h*CD44) cDNA was a gift of Dr. James B. McCarthy at the University of Minnesota (Minneapolis, MN). Rosetta 2(DE3)pLysS competent cells were purchased from EMD Biosciences (Gibbstown, NJ). HiPrep 16/60 Sephacryl S-100 HR was obtained from GE Healthcare (Piscataway, NJ). High molecular weight hyaluronan (HMW-HA) sodium salt from rooster comb was obtained from Sigma-Aldrich (St. Louis, MO). CM5 and SA Biacore Sensor Chips were obtained from GE Healthcare (Piscataway, NJ). L-Arginine, reduced/oxidized glutathione, coupling reagents, and other general buffer reagents were from Sigma-Aldrich (St. Louis, MO). Hydrazide biotin linker for making biotinylated HA was from Thermo Scientific (Rockford, IL).

6.2.2 *h*CD44 HABD Cloning, Expression and Purification

*h*CD44 HABD was cloned, expressed, and purified as previously described in Chapter 4.[88]

6.2.3 Immobilized *h*CD44 HABD SPR Assay

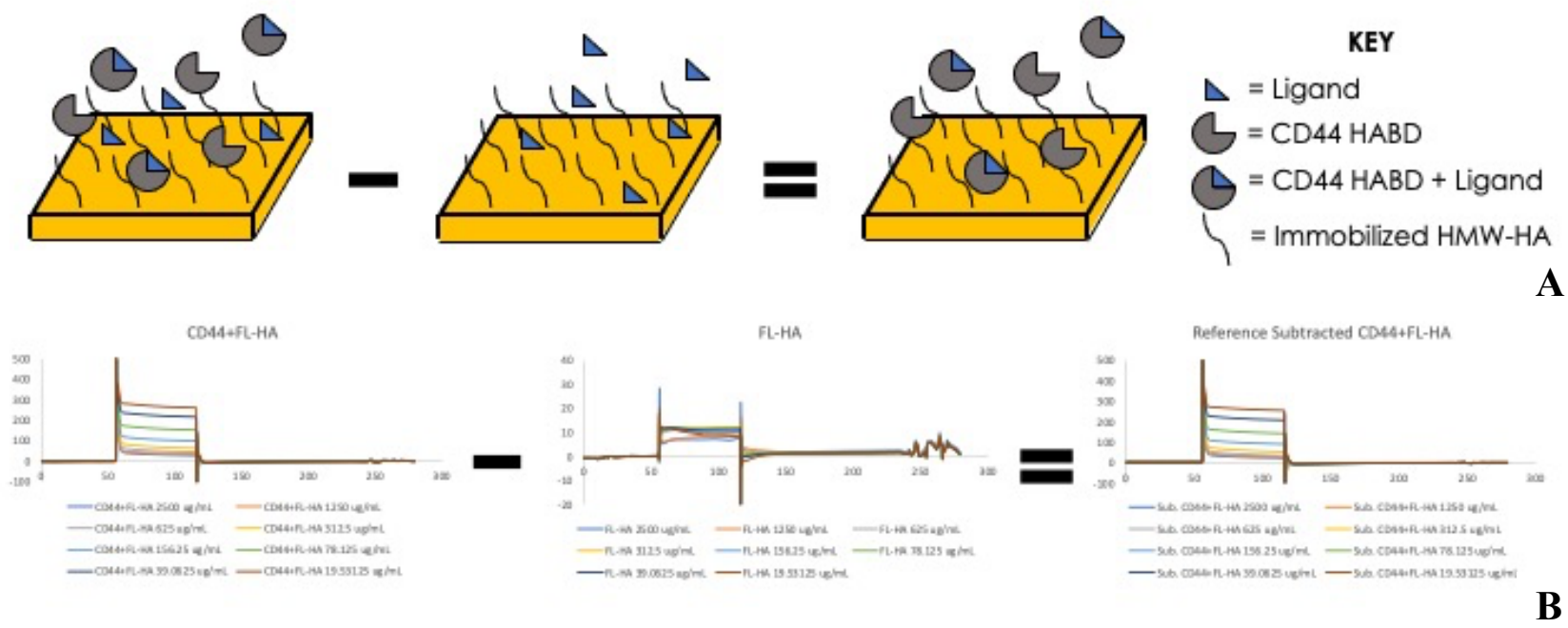
Molecular affinity for the *h*CD44 HABD was assessed by monitoring binding to a surface with covalently immobilized *h*CD44 HABD using Surface Plasmon Resonance (SPR) on a Biacore S200 (GE Healthcare) as previously described.[88] In short, a CM5 Biacore Sensor Chip surface was activated with 0.4 M 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride and 0.1 M *N*-hydroxysuccinimide. *h*CD44 HABD dialyzed against 10 mM MES pH 4.4 was flowed over the surface and bound via amine coupling until saturation at 3500 Response Units (RU) was achieved. Remaining carboxymethyl groups were blocked with 1 M ethanolamine. To quantify binding of molecules of interest, full-length HA (FL-HA) and HA-derivatives in a PBS buffer containing 0.005% Surfactant P-20 were passed over the chip surface in a two-fold dilution series (2.5 mg/mL – 19.5 µg/mL) at 30 µL/min. Equilibrium response units were measured after 60 seconds. Initial dissociation periods were set for 60 seconds but increased to 180 seconds to attempt to return binding levels back to baseline when slow dissociation was observed.

6.2.4 Immobilized HA SPR Assay with Ligand Only Binding Subtraction

HMW-HA was biotinylated using hydrazide biotin as previously described,[242] and immobilized onto a streptavidin SA Biacore Sensor Chip using established methods.[88] In short, biotinylated-HA in HEPES-EP was passed over the chip surface until the response signal stabilized at 115 RU as monitored with the Biacore S200 (GE Healthcare). For the competition-based assay, a fixed concentration of *h*CD44 HABD (~17 µM) was incubated with serial dilutions (2.5 mg/mL – 19.5 µM) of ligands (FL-HA and HA-derivatives) solubilized in HEPES-EP and placed on a plate shaker for 10 minutes at room temperature prior to placing samples on the Biacore S200. Mixtures were injected over the chip surface at 30 µL/min for 1 minute followed by 2 minutes of a dissociation phase using HEPES-EP buffer. RU values taken 5 seconds prior to the end of the equilibrium phase were plotted against concentration to calculate IC_{50} values. Data was fitted using a nonlinear regression with log[inhibitor] vs. response using GraphPad Prism.[218] Issues

surrounding non-specific binding, binding to reference chip surface, and fluidics problems of the polymeric molecules complicated analysis.

A reference subtraction experiment was completed to decrease fluidic problems and non-specific interactions associated with the polymeric nature of FL-HA and HA-derivatives. Ligand only was passed over the chip surface using the same two-fold dilution series maintaining the same buffer and experimental conditions and the resulting sensorgram was subtracted from the resulting sensorgram of CD44 preincubated with ligand. A graphical representation of the reference subtraction is shown in Scheme 6.2.



Scheme 6.2 Ligand Reference Subtracted Surface Plasmon Resonance Assay

A) Graphical representation of the reference subtraction by subtracting the effects of free ligand in solution from *h*CD44 HABD incubated with ligand. B) Spectrographs of *h*CD44 incubated with ligand, ligand only, and the corresponding reference subtraction.

6.2.5 Probability Function to Assess Distance Between Modifications

Equation 6.1 shows the probability of finding a given distance between modifications within the HA-derivatives. The equation makes the following assumptions: 1) That all molecules of HA in solution are of the same length, 2) Steric hindrance or intramolecular interactions do not impact the location of modifications, 3) The probability of finding a modification on any given glucuronic acid is equal to the loading percentage, and 4) The formula is a function of *D*, the distance between glucuronic acid monomer units. If *D* is 1 there is only one *N*-acetyl glucosamine between each modification.

Equation 6.1 Probability Function of Distances Between Modifications

$$P(D) = \frac{M}{L} \times \left[\frac{M_{-N} + D_{max} - D}{L_{-N}}\right]_N !$$

Definition of variables used in Equation 6.1:

L = number of glucuronic acids in a given length of HA

LP = loading percentage

M = number of modifications = L x LP

Dmax = maximum distance modifications = L-M+1

D = distance between modifications

P (D) = probability of finding a particular distance between modifications

N = integer of possibilities from 1 to M-1

6.3 Results and Discussion

SPR has been used previously to characterize binding involving the CD44 HABD in two different ways[88]: the Immobilized *h*CD44 HABD Assay has proved an effective tool for measuring the affinity (K_D) of small molecules for the HABD, while the Immobilized HA Assay has been used to quantify the binding of the CD44 HABD to a surface coated with HA. When that CD44

HABD is preincubated with antagonists that compete for binding with HA, the latter assay has been used to observe a reduction in CD44 HABD association to the HA surface. By observing dose-dependent responses to this competition, an IC_{50} (the concentration at which 50% of the binding of CD44 to the surface is eliminated) can be determined. It was hoped that one of these methods could be applicable in assessing the effect of chemical modifications to polymeric HA on CD44 HABD binding.

6.3.1 Inability to Quantify Binding Affinity in Immobilized HABD SPR Assay

Since all modified HA analogs are created by chemical modification of FL-HA, we first sought to characterize its affinity for the *h*CD44 HABD using the Immobilized *h*CD44 HABD SPR assay. In this assay, the ligand (e.g., FL-HA) is flowed over a surface coated with *h*CD44 HABD, and the mass change upon binding triggers the resonance effect measured in response units (RUs). This assay has been successfully applied to determine K_D of binding of both small molecule fragments, and short oligosaccharides with affinity for *h*CD44 HABD[88]. To confirm surface chip integrity, positive control, HA-10 (an HA 10-mer), was passed over the chip surface in a two-fold dilution series (250 μM – 1.95 μM). FL-HA, and in fact all the HA-derivatives (not shown) behaves very differently than small molecules in this assay, likely due to fluidic properties arising from the polymeric nature of FL-HA and the potential for high avidity with multiple (but unnumbered) sites for interaction with immobilized HABD. FL-HA was slow to reach equilibrium binding, particularly at low concentration, and washed off the surface even more slowly (Figure 6.1). Moreover, very high binding to the unmodified dextran surface of the CM5 chip was observed in instrument reference channels that further complicated data analysis. Taken together, these factors prevented an accurate K_D determination using this assay.

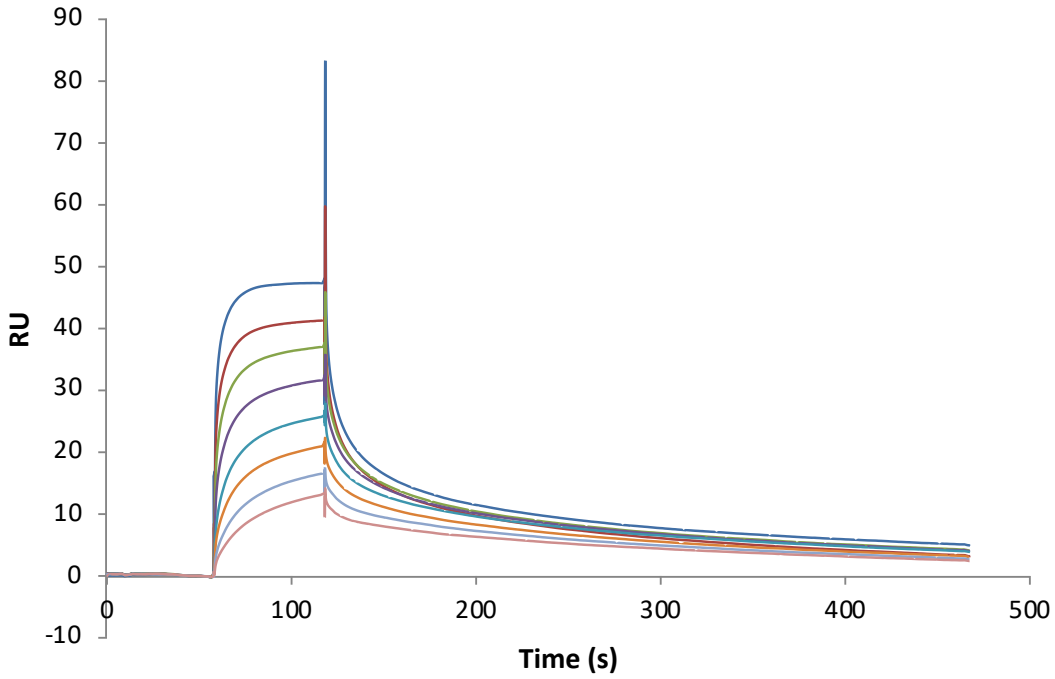


Figure 6.1 Spectrograph of FL-HA in the Immobilized *h*CD44 SPR assay

Spectrograph shows a non-plateau during the equilibrium phase and slow dissociation phase that never reaches baseline which both exemplify the non-specific interactions and avidity issue of the polymeric-HA.

6.3.2 Use of Reference Subtracted SPR Sensorgrams to Accurately Determine Effects of Modified-HA on CD44

Attempts to exploit the Immobilized HA assay exposed similar problems. In this assay, *h*CD44 HABD is flowed over an HA coated surface to which it binds, and a reduction in surface response can be observed when the HABD is preincubated with antagonists that inhibit binding through its interactions with the HABD. To determine the integrity of the chip surface various concentrations of *h*CD44 (0-150 μM) in HEPES-EP buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% (v/v) Surfactant P-20) were passed over the HMW-HA surface. An apparent K_D can be calculated for the affinity of *h*CD44 to the chip surface and previous work has shown that small molecules can specifically affect the apparent K_D.[88] Small molecules are able to compete off HMW-HA and this decrease in *h*CD44 interaction with the chip surface can be used to calculate the IC_{50} of specific small molecules. To confirm this competition, a positive control, HA-10 in a two-fold dilution series (250 μM – 1.95 μM) was incubated with a fixed concentration of *h*CD44 (~17 μM) and an IC_{50} value of 32 μM was determined for HA-10 that was close to previously published value of 48 μM. Upon introducing the polymeric FL-HA to the flow, the same problems with fluid control, binding kinetics, reference surface affinity and avidity arise to complicate both data collection and analysis.

Table 6.3 Loading Percentage of Modified-HA Compounds Calculated by NMR

Compound	Loading Percentage (%) As Calculated by NMR
A2a-25%	17
A2a-50%	30
F2a-25%	12
F2a-50%	31
G2a-25%	12
G2a-50%	23

Preliminary experiments with the HA coated surface clearly revealed differences in binding effects of FL-HA and some of its derivatives, so a reference subtraction strategy was devised to mask these effects while isolating CD44 binding. Three unique modifications to FL-HA (A2a, F2a, and G2a) at two different loading percentages dependent on the amount of aldehyde present in the Ugi reaction (25%: 1:1:0.25:1 (acid:amine:aldehyde:isocynanide) and 50%: 1:1:0.5:1) were tested in the Immobilized HA assay (Table 6.3). For simplicity the molecules will be referred to using the following syntax: Modification_aldehyde ratio (e.g., A2a_25%). The loading percentage, amount of modified glucuronic acids, as calculated by SPR is shown in Table 6.3. For FL-HA and each of the derivatives, a complete set of sensorgrams were collected for both a two-fold dilution series without any *h*CD44 HABD, and an identical series including preincubation with *h*CD44 (~17 μM). Reference subtracted sensorgrams reflecting CD44 binding to the surface were generated by subtracting curves obtained for the ligand only from those that included CD44, and IC_{50} values were computed from each reference-subtracted concentration series. A relative percent inhibition calculated by taking the IC_{50} of FL-HA and dividing it by the corresponding IC_{50} value of the HA-derivatives was calculated for comparison (Figure 6.1). The inhibition percentage of FL-HA is set at 100% and compounds observing a relative inhibition percentage greater than 100% bind tighter to *h*CD44 HABD than FL-HA. F2a-25% and G2a-25% were found to have greater inhibitory potential than FL-HA as they bind tighter to the HABD while A2a-25% was found to have lower inhibitory potential as shown in Figure 6.2. These results align well with the data from the ELISA experiments from MSU.

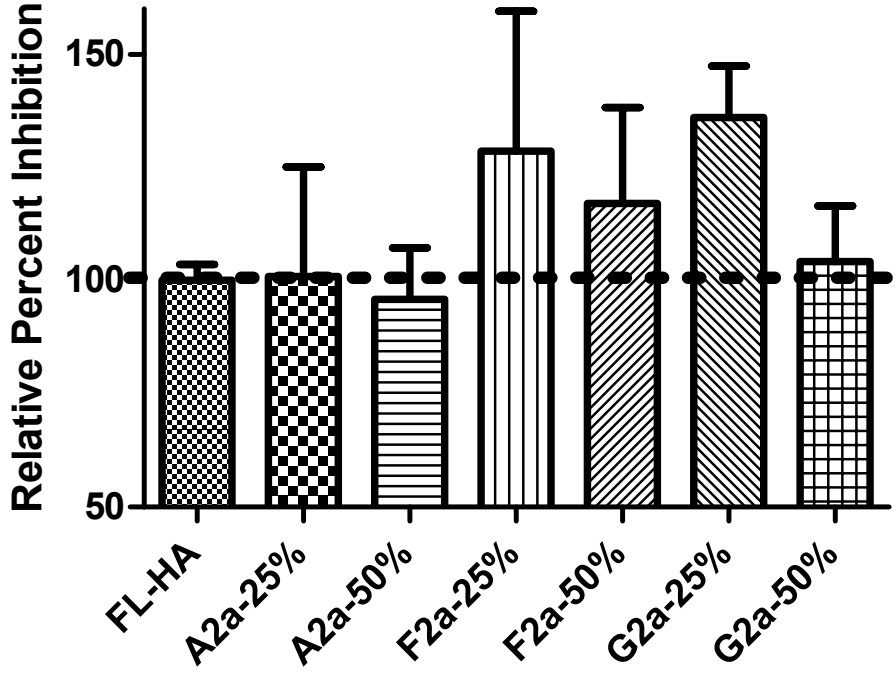


Figure 6.2 Relative Percent Inhibition of CD44

Relative inhibition of each compound at the labeled loading percentage calculated by dividing the calculated IC_{50} value of FL-HA by the IC_{50} of the modified compounds. Compounds with greater inhibitory potential have a relative percent inhibition greater than 100%. 100% inhibition is denoted with the dashed line. The error bars depict the SD of the observations collected (N=3).

6.3.3 Probability Function to Explain Effect of Loading Percentage on Inhibitory Potential

The percent inhibition was measured for A2a, F2a and G2a with samples prepared with 2 different loading percentages. While results with A2a-50% confirm that this analog has little effect on binding, both F2a and G2a show a decreased percent inhibition with higher loading percentage. This result was somewhat unexpected. If something binds, then more of something might be expected to bind better. An explanation of this may lie in statistics, however. Previous structure biology work has shown that the CD44 HABD is able to accommodate four repeating units of *N*-acetyl glucosamine and glucuronic acid (the monomeric units of FL-HA) within an extended

binding groove.[85] A plausible explanation for the decreased affinity at higher loading percentage is that the groove is only able to accommodate at most a single modification at a time. To assess the difference in binding affinity observed as a result of loading percentage a probability formula was employed. The formula is able to calculate the likelihood of finding a specific distance between modifications, D, where D equal to 1 would mean there is one N-acetyl glucosamine between modifications. The formula relies on the following assumptions: 1. All molecules of FL-HA are the same length; 2. Steric hindrance and intramolecular interactions do not impact the locations of modifications (Substitutent incorporation is truly random); 3. The probability that a specific glucuronic acid moiety is modified is equal to the loading percentage. Taking the sum of the probabilities of $D > 4$ will give the percentage of modifications that would lie further apart than the length of the extended binding groove of CD44 HABD for a given length of HA. Multiplying the number of modifications, which is equal to the product of length and loading percentage, by the percent of modifications that occur when $D>4$ will give the number of modifications that occur greater than 8 saccharide units apart. Figure 6.3 depicts the effects of loading percentage on the number of modifications that occur greater than $D=4$ and the effect of loading percentage on relative percent inhibition. Fitting the scatterplot of number of modifications that lie greater than 8 saccharide units apart with a polynomial curve fitting and solving the equation to find the local maximum, an optimal loading percentage of 19% was calculated.

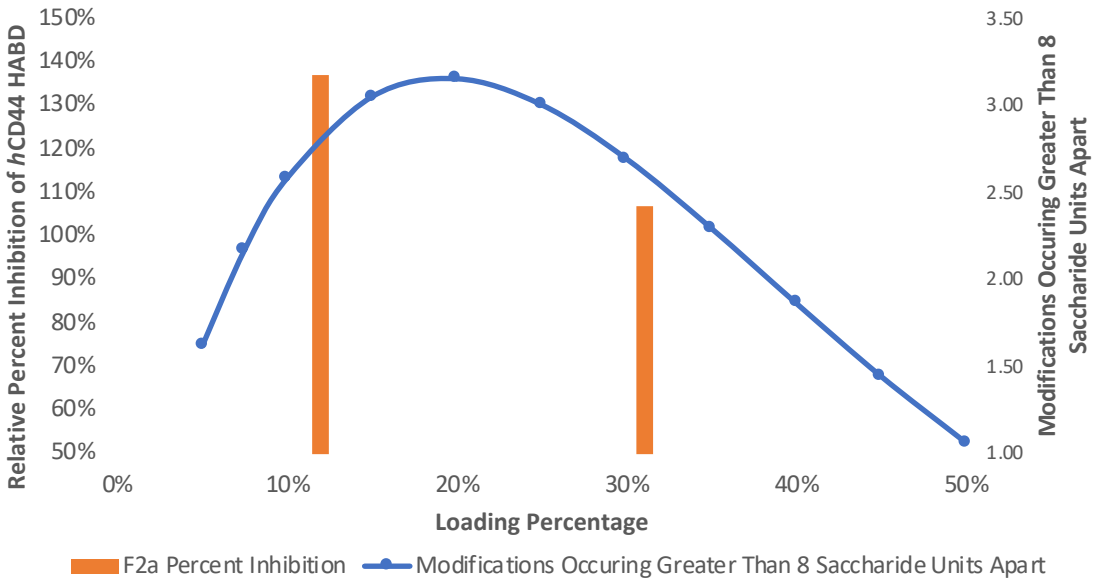


Figure 6.3 Effect of Loading Percentage on Percent Inhibition of HA-Derivatives Relative to Unmodified FL-HA

The left axis shows relative percent inhibition of *h*CD44 HABD. The bar on the left represents F2_25% and that on the right shows F2a_50%. The righthand axis represents the number of modifications that occur greater than 8 saccharide units apart. The smooth-line scatter plot represents the effects of loading percentage on the number of modifications occurring greater than 8 saccharide units apart.

6.4 Conclusion

As suspected, an accurate binding affinity, K_D, could not be determined due to the inability to quantify the molar concentrations of the modified-HA molecules and the flow and avidity issues encountered due to the polymeric nature of HA. Similar issues with flow were observed in the immobilized-HA SPR assay, but through the implementation of a ligand reference subtraction step (flowing ligand in the absence of protein over the chip surface and subtracting the resulting sensorgrams from the ligand plus protein sensorgrams) the noise related to flow could be decreased thus allowing for proper analysis of the effects of the molecules on CD44-HA interaction. Our results confirmed the ELISA data generated at Michigan State University in an orthogonal assay,

but we were able to build upon their findings by determining the effects of loading percentage on inhibition. As the loading percentage is increased, the associated inhibitory potential decreases which is contradictory to traditional small-molecule inhibition. By generating a probability function for the distance between modifications dependent on the loading percentage and length of HA, an optimal loading percentage was determined. The HABD may only be able to accommodate one modification within the groove and more than one modification within the groove causes a decrease in the relative inhibitory potential. Our studies suggest that an optimal loading percentage of 19% should give the highest inhibitory potential for any given modification of HA.

Bibliography

(1) Rang, H. P. The Receptor Concept: Pharmacology's Big Idea. *Br. J. Pharmacol.* **2006**, *147* (Suppl 1), S9. https://doi.org/10.1038/SJ.BJP.0706457.

(2) Holmes, O. W. *Currents and Counter-Currents in Medical Science*; Palala Press, 1860.

(3) Langley, J. N. On the Reaction of Cells and of Nerve-Endings to Certain Poisons, Chiefly as Regards the Reaction of Striated Muscle to Nicotine and to Curari. *J. Physiol.* **1905**, *33* (4–5), 374. https://doi.org/10.1113/JPHYSIOL.1905.SP001128.

(4) MAEHLE, D. med. A.-H. "Receptive Substances": John Newport Langley (1852–1925) and His Path to a Receptor Theory of Drug Action. *Med. Hist.* **2004**, *48* (2), 153. https://doi.org/10.1017/S0025727300000090.

(5) AV, H. The Combinations of Haemoglobin with Oxygen and with Carbon Monoxide. *Biochem. J.* **1913**, *7* (5), 471–480. https://doi.org/10.1042/BJ0070471.

(6) Monod, J.; Wyman, J.; Changeux, J. P. On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* **1965**, *12* (1), 88–118. https://doi.org/10.1016/S0022-2836(65)80285-6.

(7) AH, M.; CR, P.; RF, H. The Emergence of the Drug Receptor Theory. *Nat. Rev. Drug Discov.* **2002**, *1* (8), 637–641. https://doi.org/10.1038/NRD875.

(8) Prüll, C.-R. Part of a Scientific Master Plan? Paul Ehrlich and the Origins of His Receptor Concept. *Med. Hist.* **2003**, *47* (3), 332.

(9) Bynum, W. F. *Science and the Practice of Medicine in the Nineteenth Century*; Cambridge University Press: Cambridge, 1994.

(10) Bud, R. *Penicillin : Triumph and Tragedy*; Oxford University Press, 2007.

(11) Clark, A. J. *The Mode of Action of Drugs on Cells*; William and Wilkins Company: Baltimore, 1933. https://doi.org/10.1002/JPS.3080220628.

(12) RÖNTGEN, W. C. ON A NEW KIND OF RAYS. *Science (80-.).* **1896**, *3* (59), 227–231.

https://doi.org/10.1126/SCIENCE.3.59.227.

(13) RÖNTGEN, W. C. A NEW FORM OF RADIATION. *Science (80-.).* **1896**, *3* (72), 726–729. https://doi.org/10.1126/SCIENCE.3.72.726.

(14) Jaskolski, M.; Dauter, Z.; Wlodawer, A. A Brief History of Macromolecular Crystallography, Illustrated by a Family Tree and Its Nobel Fruits. *FEBS J.* **2014**, *281* (18), 3985–4009. https://doi.org/10.1111/FEBS.12796.

(15) Bragg, W. L. Bragg: The Diffraction of Short Electromagnetic Waves... - Google Scholar. *Proc. Camb. Philol. Soc.* **1913**.

(16) BRAGG, W. H. X-Rays and Crystals. *Nat. 1912 902243* **1912**, *90* (2243), 219–219. https://doi.org/10.1038/090219a0.

(17) Bragg, W. H.; Bragg Apr, W. L.; H Bragg, B. W.; Professor of Physics, C. The Reflection of X-Rays by Crystals. *Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character* **1913**, *88* (605), 428–438. https://doi.org/10.1098/RSPA.1913.0040.

(18) Bragg, W. H. The Reflection of X-Rays by Crystals. (II.). *Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character* **1913**, *89* (610), 246–248. https://doi.org/10.1098/RSPA.1913.0082.

(19) Bragg, M. W. L.; Ragg, W. L. B.; Communicated, B. A. (; Bragg, P. W. H.; Friedrich, * W; Snippin, P.; Laue, M. The Structure of Some Crystals as Indicated by Their Diffraction of X-Rays. *Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character* **1913**, *89* (610), 248–277. https://doi.org/10.1098/RSPA.1913.0083.

(20) Ragg, W. H. B. The Structure of the Diamond. *Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character* **1913**, *89* (610), 277–291. https://doi.org/10.1098/RSPA.1913.0084.

(21) BERNAL, J. D.; CROWFOOT, D. X-Ray Photographs of Crystalline Pepsin. *Nat. 1934 1333369* **1934**, *133* (3369), 794–795. https://doi.org/10.1038/133794b0.

(22) BERNAL, J. D. Structure of Proteins*. *Nat. 1939 1433625* **1939**, *143* (3625), 663–667.

https://doi.org/10.1038/143663a0.

(23) Crowfoot, D.; Crowfoot; Dorothy. X-Ray Single Crystal Photographs of Insulin. *Natur* **1935**, *135* (3415), 591–592. https://doi.org/10.1038/135591A0.

(24) Crowfoot, D.; Bunn, C.; Rogers-Low, B.; Turner-Jones, A. The X-Ray Crystallographic Investigation of the Structure of Penicillin. *Chem. Penicillin* **1949**, 310–367.

(25) Crowfoot Hodgkin, D. X-Ray Analysis and the Structure of Vitamin B<Subscript>12</Subscript>. *Fortschritte der Chemie Org. Naturstoffe / Prog. Chem. Org. Nat. Prod. / Progrès dans la Chim. des Subst. Org. Nat.* **1958**, 167–220. https://doi.org/10.1007/978-3-7091-7162-2_4.

(26) JC, K.; G, B.; HM, D.; RG, P.; H, W.; DC, P. A Three-Dimensional Model of the Myoglobin Molecule Obtained by x-Ray Analysis. *Nature* **1958**, *181* (4610), 662–666. https://doi.org/10.1038/181662A0.

(27) Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. C. T. Structure of Haemoglobin: A Three-Dimensional Fourier Synthesis at 5.5 Å Resolution, Obtained by X-Ray Analysis. *Nature* **1960**, *185* (4711), 416–422.

(28) Green, D. W.; Ingram, V. M.; Perutz, M. F. The Structure of Haemoglobin. IV. Sign Determination by the Isomorphous Replacement Method. *Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character* **1954**, *225*, 287–307.

(29) Adams, M. J.; Ford, G. C.; Koekoek, R.; Lentz, P. J.; McPherson, A.; Rossmann, M. G.; Smiley, I. E.; Schevitz, R. W.; Wonacott, A. J. Structure of Lactate Dehydrogenase at 2.8 Å Resolution. *Nature* **1970**, *227* (5263), 1098–1103.

(30) Matthews, B. W.; Jansonius, J. N.; Colman, P. M.; Schoenborn, B. P.; Dupourque, D. Three-Dimensional Structure of Thermolysin. *Nat. New Biol.* **1972**, *238* (80), 37–41.

(31) Dijkstra, B. W.; Drenth, J.; Kalk, K. H.; Vandermaelen, P. J. Three-Dimensional Structure and Disulfide Bond Connections in Bovine Pancreatic Phospholipase A2. *J. Mol. Biol.* **1978**, *124* (1), 53–60.

(32) Matthews, D. A.; Alden, R. A.; Bolin, J. T.; Filman, D. J.; Freer, S. T.; Hamlin, R.; Hol, W. G.; Kisliuk, R. L.; Pastore, E. J.; Plante, L. T.; Xuong, N.; Kraut, J. Dihydrofolate Reductase from Lactobacillus Casei. X-Ray Structure of the Enzyme Methotrexate.NADPH Complex. *J. Biol. Chem.* **1978**, *253* (19), 6946–6954.

(33) Rossmann, M. G. The Beginnings of Structural Biology. *Protein Sci.* **1994**, *3* (10), 1731–1733.

(34) Peters, D. C.; Noble, S. Aprotinin. *Drugs* **2012**, *57* (2), 233–260.

(35) Rühlmann, A.; Kukla, D.; Schwager, P.; Bartels, K.; Huber, R. Structure of the Complex Formed by Bovine Trypsin and Bovine Pancreatic Trypsin Inhibitor. *J. Mol. Biol.* **1973**, *77* (3), 417–436.

(36) CR, B.; PJ, G.; FE, N.; S, W.; R, W. Compounds Designed to Fit a Site of Known Structure in Human Haemoglobin. *Br. J. Pharmacol.* **1976**, *57* (2), 201–209. https://doi.org/10.1111/J.1476-5381.1976.TB07468.X.

(37) Hol, W. G. J. Protein Crystallography and Computer Graphics—toward Rational Drug Design. *Angew. Chemie Int. Ed. English* **1986**, *25* (9), 767–778. https://doi.org/10.1002/ANIE.198607673.

(38) Maveyraud, L.; Mourey, L. Protein X-Ray Crystallography and Drug Discovery. *Molecules* **2020**, *25* (5). https://doi.org/10.3390/MOLECULES25051030.

(39) Hartsuck, J. A.; Lipscomb, W. N. 1 Carboxypeptidase A. In *The Enzymes*; Boyer, P. D., Ed.; Academic Press, 1971; pp 1–56.

(40) *Foye's Principles of Medicinal Chemistry*, 6th ed.; Foye, W. O., Lemke, T. L., Williams, D. A., Eds.; Lippincott Williams and Wilkins: Philadelphia, 2008.

(41) Cushman, D. W.; Cheung, H. S.; Sabo, E. F.; Ondetti, M. A. Design of Potent Competitive Inhibitors of Angiotensin-Converting Enzyme. Carboxyalkanoyl and Mercaptoalkanoyl Amino Acids. *Biochemistry* **1977**, *16* (25), 5484–5491.

(42) Congreve, M.; Murray, C. W.; Blundell, T. L. Keynote Review: Structural Biology and

Drug Discovery. *Drug Discov. Today* **2005**, *10* (13), 895–907.

(43) Cancer Statistics.

(44) Arvelo, F.; Sojo, F.; Cotte, C. Tumor Progression and Metastasis. *Ecancermedicalscience* **2016**, *10* (617). https://doi.org/10.3332/ecancer.2016.617.

(45) Frantz, C.; Stewart, K. M.; Weaver, V. M. The Extracellular Matrix at a Glance. *J. Cell Sci.* **2010**, *123* (24), 4195–4200. https://doi.org/10.1242/JCS.023820.

(46) H, J.; A, S.; M, K.; TN, W.; R, P. Extracellular Matrix Molecules: Potential Targets in Pharmacotherapy. *Pharmacol. Rev.* **2009**, *61* (2), 198–223. https://doi.org/10.1124/PR.109.001289.

(47) L, S.; RM, S. Proteoglycans: From Structural Compounds to Signaling Molecules. *Cell Tissue Res.* **2010**, *339* (1), 237–246. https://doi.org/10.1007/S00441-009-0821-Y.

(48) Egeblad, M.; Rasch, M. G.; Weaver, V. M. Dynamic Interplay between the Collagen Scaffold and Tumor Evolution. *Curr. Opin. Cell Biol.* **2010**, *22* (5), 697–706. https://doi.org/10.1016/J.CEB.2010.08.015.

(49) Norton, W. H. J.; Ledin, J.; Grandel, H.; Neumann, C. J. HSPG Synthesis by Zebrafish Ext2 and Extl3 Is Required for Fgf10 Signalling during Limb Development. *Development* **2005**, *132* (22), 4963–4973. https://doi.org/10.1242/DEV.02084.

(50) A, P.-M.; AJ, E.; Z, W. Matrix Metalloproteinases and the Regulation of Tissue Remodelling. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (3), 221–233. https://doi.org/10.1038/NRM2125.

(51) Aitken, K. J.; Bägli, D. J. The Bladder Extracellular Matrix. Part I: Architecture, Development and Disease. *Nat. Rev. Urol. 2009 611* **2009**, *6* (11), 596–611. https://doi.org/10.1038/nrurol.2009.201.

(52) Lu, P.; Weaver, V. M.; Werb, Z. The Extracellular Matrix: A Dynamic Niche in Cancer Progression. *J. Cell Biol.* **2012**, *196* (4), 395. https://doi.org/10.1083/JCB.201102147.

(53) Bhowmick, N. A.; Neilson, E. G.; Moses, H. L. Stromal Fibroblasts in Cancer Initiation

and Progression. *Nature* **2004**, *432* (7015), 332–337.

(54) Sohara, Y.; Ishiguro, N.; Machida, K.; Kurata, H.; Thant, A. A.; Senga, T.; Matsuda, S.; Kimata, K.; Iwata, H.; Hamaguchi, M. Hyaluronan Activates Cell Motility of V-Src-Transformed Cells via Ras-Mitogen–Activated Protein Kinase and Phosphoinositide 3-Kinase-Akt in a Tumor-Specific Manner. *https://doi.org/10.1091/mbc.12.6.1859* **2017**, *12* (6), 1859–1868. https://doi.org/10.1091/MBC.12.6.1859.

(55) Collis, L.; Hall, C.; Lange, L.; Ziebell, M.; Prestwich, R.; Turley, E. A. Rapid Hyaluronan Uptake Is Associated with Enhanced Motility: Implications for an Intracellular Mode of Action. *FEBS Lett.* **1998**, *440* (3), 444–449. https://doi.org/10.1016/S0014-5793(98)01505-1.

(56) Laurent, T. C.; Laurent, U. B. G.; Fraser, J. R. E. The Structure and Function of Hyaluronan: An Overview. *Immunol. Cell Biol.* **1996**, *74* (2), a1–a7. https://doi.org/10.1038/icb.1996.32.

(57) Weigel, P. H.; Hascall, V. C.; Tammi, M. Hyaluronan Synthases *. *J. Biol. Chem.* **1997**, *272* (22), 13997–14000. https://doi.org/10.1074/JBC.272.22.13997.

(58) Karalis, T. T.; Heldin, P.; Vynios, D. H.; Neill, T.; Buraschi, S.; Iozzo, R. V.; Karamanos, N. K.; Skandalis, S. S. Tumor-Suppressive Functions of 4-MU on Breast Cancer Cells of Different ER Status: Regulation of Hyaluronan/HAS2/CD44 and Specific Matrix Effectors. *Matrix Biol.* **2019**, *78–79*, 118–138. https://doi.org/10.1016/J.MATBIO.2018.04.007.

(59) Misra, S.; Heldin, P.; Hascall, V. C.; Karamanos, N. K.; Skandalis, S. S.; Markwald, R. R.; Ghatak, S. Hyaluronan–CD44 Interactions as Potential Targets for Cancer Therapy. *FEBS J.* **2011**, *278* (9), 1429–1443. https://doi.org/10.1111/J.1742-4658.2011.08071.X.

(60) Meyer, K.; Palmer, J. W. THE POLYSACCHARIDE OF THE VITREOUS HUMOR. **1934**. https://doi.org/10.1016/S0021-9258(18)75338-6.

(61) Gandhi, N. S.; Mancera, R. L. The Structure of Glycosaminoglycans and Their

Interactions with Proteins. *Chem. Biol. Drug Des.* **2008**, *72* (6), 455–482. https://doi.org/10.1111/J.1747-0285.2008.00741.X.

(62) Laurent, T. C.; Fraser, J. R. E. Hyaluronan. *FASEB J.* **1992**, *6* (7), 2397–2404. https://doi.org/10.1096/FASEBJ.6.7.1563592.

(63) Toole, B. P. Hyaluronan: From Extracellular Glue to Pericellular Cue. *Nat. Rev. Cancer 2004 47* **2004**, *4* (7), 528–539. https://doi.org/10.1038/NRC1391.

(64) Fox, S. B.; Fawcett, J.; Jackson, D. G.; Collins, I.; Gatter, K. C.; Harris, A. L.; Gearing, A.; Simmons, D. L. Normal Human Tissues, in Addition to Some Tumors, Express Multiple Different CD44 Isoforms. *Cancer Res.* **1994**, *54* (16).

(65) Knudson, W.; Chow, G.; Knudson, C. B. CD44-Mediated Uptake and Degradation of Hyaluronan. *Matrix Biol.* **2002**, *21* (1), 15–23. https://doi.org/10.1016/S0945-053X(01)00186-X.

(66) JACKSON, D. G. Biology of the Lymphatic Marker LYVE-1 and Applications in Research into Lymphatic Trafficking and Lymphangiogenesis. *APMIS* **2004**, *112* (7–8), 526–538. https://doi.org/10.1111/J.1600-0463.2004.APM11207-0811.X.

(67) and, R. S.; Mark J. Jedrzejas*, ‡. Hyaluronidases: Their Genomics, Structures, and Mechanisms of Action. *Chem. Rev.* **2006**, *106* (3), 818–839. https://doi.org/10.1021/CR050247K.

(68) Roden, L.; Campbell, P.; Fraser, R.; Laurent, T.; Pertoft, H.; Thompson, J. Enzymic Pathways of Hyaluronan Catabolism. In *The Biology of Hyaluronan*; Wiley, 1989; pp 60–86.

(69) Uchiyama, H.; Dobashi, Y.; Ohkouchi, K.; Nagasawas, K. Chemical Change Involved in the Oxidative Reductive Depolymerization of Hyaluronic Acid*. *J. BIOLOOICAL Chem.* **1990**, *265* (14), 7753–7759. https://doi.org/10.1016/S0021-9258(19)38993-8.

(70) TB, C.; GI, F.; R, S. Hyaluronidases in Tissue Invasion. *Invasion Metastasis* **1997**, *17* (6), 297–311.

(71) R, S.; AA, A.; KN, S. Hyaluronan Fragments: An Information-Rich System. *Eur. J. Cell Biol.* **2006**, *85* (8), 699–715. https://doi.org/10.1016/J.EJCB.2006.05.009.

(72) Y, T.; L, L.; M, K.; T, A.; A, M.; H, Y.; P, H. Hyaluronan Fragments Induce Endothelial Cell Differentiation in a CD44- and CXCL1/GRO1-Dependent Manner. *J. Biol. Chem.* **2005**, *280* (25), 24195–24204. https://doi.org/10.1074/JBC.M411913200.

(73) S, M.-N.; J, G.; S, K.; M, S. Oligosaccharides of Hyaluronan Induce Angiogenesis through Distinct CD44 and RHAMM-Mediated Signalling Pathways Involving Cdc2 and Gamma-Adducin. *Int. J. Oncol.* **2009**, *35* (4). https://doi.org/10.3892/IJO_00000389.

(74) Sugahara, K. N.; Murai, T.; Nishinakamura, H.; Kawashima, H.; Saya, H.; Miyasaka, M. Hyaluronan Oligosaccharides Induce CD44 Cleavage and Promote Cell Migration in CD44-Expressing Tumor Cells. *J. Biol. Chem.* **2003**, *278* (34), 32259–32265. https://doi.org/10.1074/jbc.M300347200.

(75) Sugahara, K. N.; Hirata, T.; Hayasaka, H.; Stern, R.; Murai, T.; Miyasaka, M. Tumor Cells Enhance Their Own CD44 Cleavage and Motility by Generating Hyaluronan Fragments. *J. Biol. Chem.* **2006**, *281* (9), 5861–5868. https://doi.org/10.1074/jbc.M506740200.

(76) EA, T.; PW, N.; LY, B. Signaling Properties of Hyaluronan Receptors. *J. Biol. Chem.* **2002**, *277* (7), 4589–4592. https://doi.org/10.1074/JBC.R100038200.

(77) Day, A. J.; Prestwich, G. D. Hyaluronan-Binding Proteins: Tying Up the Giant *. *J. Biol. Chem.* **2002**, *277* (7), 4585–4588. https://doi.org/10.1074/JBC.R100036200.

(78) Ponta, H.; Sherman, L.; Herrlich, P. A. CD44: From Adhesion Molecules to Signalling Regulators. *Nat. Rev. Mol. Cell Biol. 2003 41* **2003**, *4* (1), 33–45. https://doi.org/10.1038/nrm1004.

(79) Bourguignon, L. Y. W. CD44-Mediated Oncogenic Signaling and Cytoskeleton Activation During Mammary Tumor Progression. *J. Mammary Gland Biol. Neoplasia 2001 63* **2001**, *6* (3), 287–297. https://doi.org/10.1023/A:1011371523994.

(80) Thorne, R. F.; Legg, J. W.; Isacke, C. M. The Role of the CD44 Transmembrane and Cytoplasmic Domains in Co-Ordinating Adhesive and Signalling Events. *J. Cell Sci.* **2004**, *117* (3), 373–380. https://doi.org/10.1242/JCS.00954.

(81) Kohda, D.; Morton, C. J.; Parkar, A. A.; Campbell, I. D.; Day, A. J. Solution Structure of the Link Module: A Hyaluronan-Binding Domain Involved in Extracellular Matrix Stability and Cell Migration. *Cell* **1996**, *86*, 767–775.

(82) Day, A. J. The Structure and Regulation of Hyaluronan-Binding Proteins. *Biochem. Soc. Trans.* **1999**, *27* (2), 115–121. https://doi.org/10.1042/BST0270115.

(83) Yang, B.; Yang, B. L.; Savani, R. C.; Turley, E. A. Identification of a Common Hyaluronan Binding Motif in the Hyaluronan Binding Proteins RHAMM, CD44 and Link Protein. *EMBO J.* **1994**, *13* (2), 286–296. https://doi.org/10.1002/J.1460-2075.1994.TB06261.X.

(84) Chen, C.; Zhao, S.; Karnad, A.; Freeman, J. W. The Biology and Role of CD44 in Cancer Progression: Therapeutic Implications. *J. Hematol. Oncol. 2018 111* **2018**, *11* (1), 1–23. https://doi.org/10.1186/S13045-018-0605-5.

(85) Banerji, S.; Wright, A. J.; Noble, M.; Mahoney, D. J.; Campbell, I. D.; Day, A. J.; Jackson, D. G. Structures of the Cd44-Hyaluronan Complex Provide Insight into a Fundamental Carbohydrate-Protein Interaction. *Nat. Struct. Mol. Biol.* **2007**, *14* (3), 234–239. https://doi.org/10.1038/nsmb1201.

(86) Banerji, S.; Day, A. J.; Kahmann, J. D.; Jackson, D. G. Characterization of a Functional Hyaluronan-Binding Domain from the Human CD44 Molecule Expressed in Escherichia Coil. *Protein Expr. Purif.* **1998**, *14* (3), 371–381. https://doi.org/10.1006/prep.1998.0971.

(87) Scott, D. E.; Bayly, A. R.; Abell, C.; Skidmore, J. Small Molecules, Big Targets: Drug Discovery Faces the Protein–Protein Interaction Challenge. *Nat. Rev. Drug Discov. 2016 158* **2016**, *15* (8), 533–550. https://doi.org/10.1038/nrd.2016.29.

(88) Liu, L. K.; Finzel, B. C. Fragment-Based Identification of an Inducible Binding Site on

Cell Surface Receptor CD44 for the Design of Protein-Carbohydrate Interaction Inhibitors. *J. Med. Chem.* **2014**, *57* (6), 2714–2725. https://doi.org/10.1021/jm5000276.

(89) Lennon, F. E.; Singleton, P. A. Hyaluronan Regulation of Vascular Integrity. *Am. J. Cardiovasc. Dis.* **2011**, *1* (3), 200–213.

(90) McKee, C. M.; Penno, M. B.; Cowman, M.; Burdick, M. D.; Strieter, R. M.; Bao, C.; Noble, P. W. Hyaluronan (HA) Fragments Induce Chemokine Gene Expression in Alveolar Macrophages: The Role of HA Size and CD44. *J. Clin. Invest.* **1996**, *98* (10), 2403–2413. https://doi.org/10.1172/JCI119054.

(91) Slevin, M.; Kumar, S.; Gaffney, J. Angiogenic Oligosaccharides of Hyaluronan Induce Multiple Signaling Pathways Affecting Vascular Endothelial Cell Mitogenic and Wound Healing Responses. *J. Biol. Chem.* **2002**, *277* (43), 41046–41059. https://doi.org/10.1074/jbc.M109443200.

(92) S, G.; S, M.; BP, T.; Ghatak, S.; Misra, S.; Toole, B. P. Hyaluronan Oligosaccharides Inhibit Anchorage-Independent Growth of Tumor Cells by Suppressing the Phosphoinositide 3-Kinase/Akt Cell Survival Pathway. **2002**, *277* (41), 38013–38020.

(93) Slomiany, M. G.; Dai, L.; Bomar, P. A.; Knackstedt, T. J.; Kranc, D. A.; Tolliver, L.; Maria, B. L.; Toole, B. P. Abrogating Drug Resistance in Malignant Peripheral Nerve Sheath Tumors by Disrupting Hyaluronan-CD44 Interactions with Small Hyaluronan Oligosaccharides. *Cancer Res.* **2009**, *69* (12), 4992–4998. https://doi.org/10.1158/0008-5472.CAN-09-0143.

(94) A, Z.; R, B.; K, T.; J, B.; V, H.; D, J.-W.; M, W.; S, R.; A, H. Therapeutic Effects of Antibodies against Adhesion Molecules in Murine Collagen Type II-Induced Arthritis. *Autoimmunity* **1995**, *21* (4), 245–252. https://doi.org/10.3109/08916939509001943.

(95) K, M.; FR, B.; JH, K.; TT, G. Anti-CD44 Treatment Abrogates Tissue Oedema and Leukocyte Infiltration in Murine Arthritis. *Nat. Med.* **1995**, *1* (6), 558–563. https://doi.org/10.1038/NM0695-558.

(96) G, S.; X, L.; L, Z.; L, W.; Y, F.; ZC, H. HI44a, an Anti-CD44 Monoclonal Antibody, Induces Differentiation and Apoptosis of Human Acute Myeloid Leukemia Cells. *Leuk. Res.* **2004**, *28* (10), 1089–1096. https://doi.org/10.1016/J.LEUKRES.2004.02.005.

(97) Colnot, D. R.; Wilhelm, A. J.; Cloos, J.; Roos, J. C.; de Bree, R.; Quak, J. J.; Snow, G. B.; van Dongen, G. A. M. S. Evaluation of Limited Blood Sampling in a Preceding 99mTc-Labeled Diagnostic Study to Predict the Pharmacokinetics and Myelotoxicity of 186Re-CMAb U36 Radioimmunotherapy. *J. Nucl. Med.* **2001**, *42* (9).

(98) de Bree, R.; Roos, J. C.; Quak, J. J.; den Hollander, W.; Wilhelm, A. J.; van Lingen, A.; Snow, G. B.; Dongen, G. A. Biodistribution of Radiolabeled Monoclonal Antibody E48 IgG and F(Ab')2 in Patients with Head and Neck Cancer. *Clin. Cancer Res.* **1995**, *1* (3).

(99) de Bree, R.; Roos, J. C.; Quak, J. J.; den Hollander, W.; Snow, G. B.; van Dongen, G. A. Radioimmunoscintigraphy and Biodistribution of Technetium-99m-Labeled Monoclonal Antibody U36 in Patients with Head and Neck Cancer. *Clin. Cancer Res.* **1995**, *1* (6).

(100) DR, C.; JC, R.; R, de B.; AJ, W.; JA, K.; G, H.; KH, H.; G, S.; GB, S.; GA, van D. Safety, Biodistribution, Pharmacokinetics, and Immunogenicity of 99mTc-Labeled Humanized Monoclonal Antibody BIWA 4 (Bivatuzumab) in Patients with Squamous Cell Carcinoma of the Head and Neck. *Cancer Immunol. Immunother.* **2003**, *52* (9), 576–582. https://doi.org/10.1007/S00262-003-0396-5.

(101) BP, T.; S, G.; S, M. Hyaluronan Oligosaccharides as a Potential Anticancer Therapeutic. *Curr. Pharm. Biotechnol.* **2008**, *9* (4), 249–252. https://doi.org/10.2174/138920108785161569.

(102) H, U.; Y, N.; W, K.; CB, K.; E, A.; E, K.; N, F.; J, W.; N, I. Therapeutic Potential of Hyaluronan Oligosaccharides for Bone Metastasis of Breast Cancer. *J. Orthop. Res.* **2012**, *30* (4), 662–672. https://doi.org/10.1002/JOR.21557.

(103) K, F.; A, H.; A, S.; B, H.; JP, S.; V, O.-R. Opposing Effects of High- and Low-Molecular Weight Hyaluronan on CXCL12-Induced CXCR4 Signaling Depend on CD44. *Cell Death*

Dis. **2013**, *4* (10). https://doi.org/10.1038/CDDIS.2013.364.

(104) Wang, C.; Wang, Z.; Chen, C.; Fu, X.; Wang, J.; Fei, X.; Yan, X.; Xu, R. A Low MW Inhibitor of CD44 Dimerization for the Treatment of Glioblastoma. *Br. J. Pharmacol.* **2020**, *177* (13), 3009–3023. https://doi.org/10.1111/bph.15030.

(105) Sellés Vidal, L.; Kelly, C. L.; Mordaka, P. M.; Heap, J. T. Review of NAD(P)H-Dependent Oxidoreductases: Properties, Engineering and Application. *Biochimica et Biophysica Acta - Proteins and Proteomics*. 2018. https://doi.org/10.1016/j.bbapap.2017.11.005.

(106) Meijers, R.; Morris, R. J.; Adolph, H. W.; Merli, A.; Lamzin, V. S.; Cedergren-Zeppezauer, E. S. On the Enzymatic Activation of NADH *. *J. Biol. Chem.* **2001**, *276* (12), 9316–9321. https://doi.org/10.1074/JBC.M010870200.

(107) Holm, A. K.; Blank, L. M.; Oldiges, M.; Schmid, A.; Solem, C.; Jensen, P. R.; Vemuri, G. N. Metabolic and Transcriptional Response to Cofactor Perturbations in Escherichia Coli. *J. Biol. Chem.* **2010**, *285* (23), 17498–17506. https://doi.org/10.1074/JBC.M109.095570.

(108) Sauer, U.; Canonaco, F.; Heri, S.; Perrenoud, A.; Fischer, E. The Soluble and Membrane-Bound Transhydrogenases UdhA and PntAB Have Divergent Functions in NADPH Metabolism of Escherichia Coli*. *J. Biol. Chem.* **2004**, *279* (8), 6613–6619. https://doi.org/10.1074/JBC.M311657200.

(109) De Graef, M. R.; Alexeeva, S.; Snoep, J. L.; Teixeira De Mattos, M. J. The Steady-State Internal Redox State (NADH/NAD) Reflects the External Redox State and Is Correlated with Catabolic Adaptation in Escherichia Coli. *J. Bacteriol.* **1999**, *181* (8), 2351–2357. https://doi.org/10.1128/JB.181.8.2351-2357.1999.

(110) Auriol, C.; Bestel-Corre, G.; Claude, J.-B.; Soucaille, P.; Meynial-Salles, I. Stress-Induced Evolution of Escherichia Coli Points to Original Concepts in Respiratory Cofactor Selectivity. *Proc. Natl. Acad. Sci.* **2011**, *108* (4), 1278–1283. https://doi.org/10.1073/PNAS.1010431108.

(111) Cahn, J. K. B.; Werlang, C. A.; Baumschlager, A.; Brinkmann-Chen, S.; Mayo, S. L.; Arnold, F. H. A General Tool for Engineering the NAD/NADP Cofactor Preference of Oxidoreductases. *ACS Synth. Biol.* **2016**, *6* (2), 326–333. https://doi.org/10.1021/ACSSYNBIO.6B00188.

(112) ADAMS, M. J.; FORD, G. C.; KOEKOEK, R.; LENTZ, P. J.; McPHERSON, A.; ROSSMANN, M. G.; SMILEY, I. E.; SCHEVITZ, R. W.; WONACOTT, A. J. Structure of Lactate Dehydrogenase at 2.8 Å Resolution. *Nat. 1970 2275263* **1970**, *227* (5263), 1098–1103. https://doi.org/10.1038/2271098a0.

(113) Rossmann, M. G.; Moras, D.; Olsen, K. W. Chemical and Biological Evolution of a Nucleotide-Binding Protein. *Nat. 1974 2505463* **1974**, *250* (5463), 194–199. https://doi.org/10.1038/250194a0.

(114) Schulz, G. E.; Schirmer, R. H.; Pai, E. F. FAD-Binding Site of Glutathione Reductase. *J. Mol. Biol.* **1982**, *160* (2), 287–308. https://doi.org/10.1016/0022-2836(82)90177-2.

(115) HANUKOGLU, I.; GUTFINGER, T. CDNA Sequence of Adrenodoxin Reductase. *Eur. J. Biochem.* **1989**, *180* (2), 479–484. https://doi.org/10.1111/J.1432-1033.1989.TB14671.X.

(116) Bottoms, C. A.; Smith, P. E.; Tanner, J. J. A Structurally Conserved Water Molecule in Rossmann Dinucleotide-Binding Domains. *Protein Sci.* **2002**, *11* (9), 2125–2137. https://doi.org/10.1110/PS.0213502.

(117) Hanukoglu, I. Proteopedia: Rossmann Fold: A Beta-Alpha-Beta Fold at Dinucleotide Binding Sites. *Biochem. Mol. Biol. Educ.* **2015**, *43* (3), 206–209. https://doi.org/10.1002/BMB.20849.

(118) Minárik, P.; Tomaásková, N.; Kollárová, M.; Antalík, M. Malate Dehydrogenases - Structure and Function. *Gen. Physiol. Biophys.* **2002**.

(119) Silverstein, E.; Sulebele, G. Catalytic Mechanism of Pig Heart Mitochondrial Malate Dehydrogenase Studied by Kinetics at Equilibrium. *Biochemistry* **1969**, *8* (6), 2543–2550. https://doi.org/10.1021/BI00834A042.

(120) Lunt, S. Y.; Vander Heiden, M. G. Aerobic Glycolysis: Meeting the Metabolic Requirements of Cell Proliferation. *Annu. Rev. Cell Dev. Biol.* **2011**, *27* (1), 441–464. https://doi.org/10.1146/annurev-cellbio-092910-154237.

(121) Schneider, M.; Knuesting, J.; Birkholz, O.; Heinisch, J. J.; Scheibe, R. Cytosolic GAPDH as a Redox-Dependent Regulator of Energy Metabolism. *BMC Plant Biol.* **2018**, *18* (1). https://doi.org/10.1186/S12870-018-1390-6.

(122) Goldman, R. D.; Kaplan, N. O.; Hall, T. C. Lactic Dehydrogenase in Human Neoplastic Tissues. *Cancer Res.* **1964**, *24*, 389–399.

(123) MV, L.; JW, L. The Warburg Effect: How Does It Benefit Cancer Cells? *Trends Biochem. Sci.* **2016**, *41* (3), 211–218. https://doi.org/10.1016/J.TIBS.2015.12.001.

(124) XH, L.; MA, M.; A, K.; OA, O.; C, J.; CB, K.; R, J.; A, K. The Proapoptotic Function of Noxa in Human Leukemia Cells Is Regulated by the Kinase Cdk5 and by Glucose. *Mol. Cell* **2010**, *40* (5), 823–833. https://doi.org/10.1016/J.MOLCEL.2010.11.035.

(125) Hanse, E. A.; Ruan, C.; Kachman, M.; Wang, D.; Lowman, X. H.; Kelekar, A. Cytosolic Malate Dehydrogenase Activity Helps Support Glycolysis in Actively Proliferating Cells and Cancer. *Oncogene* **2017**, *36* (27), 3915–3924. https://doi.org/10.1038/onc.2017.36.

(126) K, B.; T, W.; WW, C.; E, F.; M, A.-R.; DM, S. An Essential Role of the Mitochondrial Electron Transport Chain in Cell Proliferation Is to Enable Aspartate Synthesis. *Cell* **2015**, *162* (3), 540–551. https://doi.org/10.1016/J.CELL.2015.07.016.

(127) E, C.; J, G.; U, D.; BE, G.; SO, S.; BA, A.; A, J.; CJ, B.; ML, H.; E, L.; Y, A.; B, R.; AP, G.; C, S.; N, S. The CBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov.* **2012**, *2* (5), 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095.

(128) J, G.; BA, A.; U, D.; G, D.; B, G.; SO, S.; Y, S.; A, J.; R, S.; E, L.; E, C.; C, S.; N, S. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the CBioPortal. *Sci. Signal.* **2013**, *6* (269). https://doi.org/10.1126/SCISIGNAL.2004088.

(129) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank . *Nucleic Acids Res.* **2000**, *28* (1), 235–242.

(130) Beaupre, B. A.; Roman, J. V.; Hoag, M. R.; Meneely, K. M.; Silvaggi, N. R.; Lamb, A. L.; Moran, G. R. Ligand Binding Phenomena That Pertain to the Metabolic Function of Renalase. *Arch. Biochem. Biophys.* **2016**, *612*, 46–56. https://doi.org/10.1016/j.abb.2016.10.011.

(131) Zaitseva, J.; Meneely, K. M.; Lamb, A. L. Structure of Escherichia Coli Malate Dehydrogenase at 1.45 Å Resolution. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2009**, *65* (9), 866–869. https://doi.org/10.1107/S1744309109032217.

(132) Hall, M. D.; Levitt, D. G.; Banaszak, L. J. Crystal Structure of Escherichia Coli Malate Dehydrogenase. A Complex of the Apoenzyme and Citrate at 1·87 Å Resolution. *J. Mol. Biol.* **1992**, *226* (3), 867–882. https://doi.org/10.1016/0022-2836(92)90637-Y.

(133) Cox, B.; Chit, M. M.; Weaver, T.; Gietl, C.; Bailey, J.; Bell, E.; Banaszak, L. Organelle and Translocatable Forms of Glyoxysomal Malate Dehydrogenase: The Effect of the N-Terminal Presequence. *FEBS J.* **2005**, *272* (3), 643–654. https://doi.org/10.1111/j.1742-4658.2004.04475.x.

(134) Gleason, W. B.; Fu, Z.; Birktoft, J.; Banaszak, L. Refined Crystal Structure of Mitochondrial Malate Dehydrogenase from Porcine Heart and the Consensus Structure for Dicarboxylic Acid Oxidoreductases. *Biochemistry* **1994**, *33* (8), 2078–2088. https://doi.org/10.1021/bi00174a014.

(135) Kelly, C. A.; Nishiyama, M.; Ohnishi, Y.; Beppu, T.; Birktoft, J. J. Determinants of Protein Thermostability Observed in the 1.9-Å Crystal Structure of Malate Dehydrogenase from the Thermophilic Bacterium Thermus Flavus. *Biochemistry* **1993**, *32* (15), 3913–3922. https://doi.org/10.1021/bi00066a010.

(136) Ferraris, D. M.; Spallek, R.; Oehlmann, W.; Singh, M.; Rizzi, M. Structures of Citrate

Synthase and Malate Dehydrogenase of Mycobacterium Tuberculosis. *Proteins Struct. Funct. Bioinforma.* **2015**, *83* (2), 389–394. https://doi.org/10.1002/prot.24743.

(137) Huang, J.; Niazi, A. K.; Young, D.; Rosado, L. A.; Vertommen, D.; Bodra, N.; Abdelgawwad, M. R.; Vignols, F.; Wei, B.; Wahni, K.; Bashandy, T.; Bariat, L.; Van Breusegem, F.; Messens, J.; Reichheld, J. P. Self-Protection of Cytosolic Malate Dehydrogenase against Oxidative Stress in Arabidopsis. *J. Exp. Bot.* **2018**, *69* (14), 3491–3505. https://doi.org/10.1093/jxb/erx396.

(138) Chapman, A. D. M.; Cortés, A.; Dafforn, T. R.; Clarke, A. R.; Brady, R. L. Structural Basis of Substrate Specificity in Malate Dehydrogenases: Crystal Structure of a Ternary Complex of Porcine Cytoplasmic Malate Dehydrogenase, α-Ketomalonate and TetrahydoNAD. *J. Mol. Biol.* **1999**, *285* (2), 703–712. https://doi.org/10.1006/jmbi.1998.2357.

(139) Rao, S. T.; Rossmann, M. G. Comparison of Super-Secondary Structures in Proteins. *J. Mol. Biol.* **1973**, *76* (2), 241–256. https://doi.org/10.1016/0022-2836(73)90388-4.

(140) DR, B.; E, R.; LJ, B. Engineering the Quaternary Structure of an Enzyme: Construction and Analysis of a Monomeric Form of Malate Dehydrogenase from Escherichia Coli. *Protein Sci.* **1994**, *3* (11), 2023–2032. https://doi.org/10.1002/PRO.5560031115.

(141) DJ, N.; J, M.; MD, S.; AR, C.; JJ, H.; T, A.; CR, G. The Importance of Arginine 102 for the Substrate Specificity of Escherichia Coli Malate Dehydrogenase. *Biochem. Biophys. Res. Commun.* **1992**, *189* (2), 1057–1062. https://doi.org/10.1016/0006-291X(92)92311-K.

(142) CR, G.; DJ, N. Malate Dehydrogenase: A Model for Structure, Evolution, and Catalysis. *Protein Sci.* **1994**, *3* (10), 1883–1888. https://doi.org/10.1002/PRO.5560031027.

(143) Lamzin, V. S.; Dauter, Z.; Wilson, K. S. Dehydrogenation through the Looking–Glass. *Nat. Struct. Biol. 1994 15* **1994**, *1* (5), 281–282. https://doi.org/10.1038/nsb0594-281.

(144) KA, J.; JR, H.; JS, S.; IR, M.; M, N.; DM, M.; DW, D. Neuropathological Background of

Phenotypical Variability in Frontotemporal Dementia. *Acta Neuropathol.* **2011**, *122* (2), 137–153. https://doi.org/10.1007/S00401-011-0839-6.

(145) T, A.; JT, S.; M, H. Tau and Tauopathies. *Brain Res. Bull.* **2016**, *126* (Pt 3), 238–292. https://doi.org/10.1016/J.BRAINRESBULL.2016.08.018.

(146) MG, S.; M, G.; RA, C.; JR, M.; MR, F.; B, G. Familial Multiple System Tauopathy with Presenile Dementia: A Disease with Abundant Neuronal and Glial Tau Filaments. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (8), 4113–4118. https://doi.org/10.1073/PNAS.94.8.4113.

(147) Josephs, K. A. Current Understanding of Neurodegenerative Diseases Associated With the Protein Tau. *Mayo Clin. Proc.* **2017**, *92* (8), 1291–1303. https://doi.org/10.1016/J.MAYOCP.2017.04.016.

(148) Alzheimer's Facts and Figures Report | Alzheimer's Association https://www.alz.org/alzheimers-dementia/facts-figures (accessed Aug 6, 2021).

(149) Horvitz, H. R. Genetic Control of Programmed Cell Death in the Nematode Caenorhabditis Elegans. *Cancer Res.* **1994**, *59* (7 SUPPL.), 1–13. https://doi.org/10.1007/978-1-4757-9217-1_1.

(150) Jacobson, M. D.; Weil, M. Programmed Cell Death Review in Animal Development. *Cell* **1997**, *88*, 347–354.

(151) JF, K.; AH, W.; AR, C. Apoptosis: A Basic Biological Phenomenon with Wide-Ranging Implications in Tissue Kinetics. *Br. J. Cancer* **1972**, *26* (4), 239–257. https://doi.org/10.1038/BJC.1972.33.

(152) AH, W.; JF, K.; AR, C. Cell Death: The Significance of Apoptosis. *Int. Rev. Cytol.* **1980**, *68* (C), 251–306. https://doi.org/10.1016/S0074-7696(08)62312-8.

(153) JE, S.; HR, H. Post-Embryonic Cell Lineages of the Nematode, Caenorhabditis Elegans. *Dev. Biol.* **1977**, *56* (1), 110–156. https://doi.org/10.1016/0012-1606(77)90158-0.

(154) HM, E.; HR, H. Genetic Control of Programmed Cell Death in the Nematode C. Elegans.

Cell **1986**, *44* (6), 817–829. https://doi.org/10.1016/0092-8674(86)90004-8.

(155) DP, C.; CJ, K.; B, M.; N, N.; K, V. N.; TA, G.; CJ, M.; SR, K.; T, D.; LA, C. Molecular Cloning of the Interleukin-1 Beta Converting Enzyme. *Science* **1992**, *256* (5053), 97–100. https://doi.org/10.1126/SCIENCE.1373520.

(156) NA, T.; HG, B.; JR, C.; KT, C.; AD, H.; MJ, K.; DK, M.; SM, M.; JR, W.; J, A. A Novel Heterodimeric Cysteine Protease Is Required for Interleukin-1 Beta Processing in Monocytes. *Nature* **1992**, *356* (6372), 768–774. https://doi.org/10.1038/356768A0.

(157) Yuan, J.; Shaham, S.; Ledoux, S.; Ellis, H. M.; Horvitz, H. R. The C. Elegans Cell Death Gene Ced-3 Encodes a Protein Similar to Mammalian Interleukin-1β-Converting Enzyme. *Cell* **1993**, *75* (4), 641–652. https://doi.org/10.1016/0092-8674(93)90485-9.

(158) Miura, M.; Zhu, H.; Rotello, R.; Hartwieg, E. A.; Yuan, J. Induction of Apoptosis in Fibroblasts by IL-1β-Converting Enzyme, a Mammalian Homolog of the C. Elegans Cell Death Gene Ced-3. *Cell* **1993**, *75* (4), 653–660. https://doi.org/10.1016/0092-8674(93)90486-A.

(159) Bennet, L.; Downs, V. Introduction to Caspases. In *Caspase-3: Structure, Functions and Interactions*; Bennet, L., Ed.; Nova Science Publishers: New York, NY, 2020; pp 1–20.

(160) Li, J.; Yuan, J. Caspases in Apoptosis and Beyond. *Oncogene 2008 2748* **2008**, *27* (48), 6194–6206. https://doi.org/10.1038/onc.2008.297.

(161) Shi, Y. Mechanisms of Caspase Activation and Inhibition during Apoptosis. *Mol. Cell* **2002**, *9* (3), 459–470. https://doi.org/10.1016/S1097-2765(02)00482-3.

(162) Fuentes-Prior, P.; Salvesen, G. S. The Protein Structures That Shape Caspase Activity, Specificity, Activation and Inhibition. *Biochem. J.* **2004**, *384* (Pt 2), 201–232. https://doi.org/10.1042/BJ20041142.

(163) Ballatore, C.; Lee, V. M.-Y.; Trojanowski, J. Q. Tau-Mediated Neurodegeneration in Alzheimer's Disease and Related Disorders. *Nat. Rev. Neurosci. 2007 89* **2007**, *8* (9), 663–672. https://doi.org/10.1038/nrn2194.

(164) G, L.; RL, N.; KS, K. The Microtubule Binding Domain of Tau Protein. *Neuron* **1989**, *2* (6), 1615–1624. https://doi.org/10.1016/0896-6273(89)90050-0.

(165) LA, A. Microtubule Structure and Its Stabilisation. *Org. Biomol. Chem.* **2004**, *2* (15), 2153–2160. https://doi.org/10.1039/B403634D.

(166) S, K.; J, F.; MJ, S.; M, G.; LA, A. Repeat Motifs of Tau Bind to the Insides of Microtubules in the Absence of Taxol. *EMBO J.* **2003**, *22* (1), 70–77. https://doi.org/10.1093/EMBOJ/CDG001.

(167) Mazanetz, M. P.; Fischer, P. M. Untangling Tau Hyperphosphorylation in Drug Design for Neurodegenerative Diseases. *Nat. Rev. Drug Discov. 2007 66* **2007**, *6* (6), 464–479. https://doi.org/10.1038/nrd2111.

(168) M, G.; JP, D.; A, D.; J, L.; R, M. Sequence of Neurofibrillary Changes in Aging and Alzheimer's Disease: A Confocal Study with Phospho-Tau Antibody, AD2. *J. Alzheimers. Dis.* **2001**, *3* (4), 417–425. https://doi.org/10.3233/JAD-2001-3409.

(169) S, M.; N, S.; Y, S.; M, M.; Y, Y.; H, K.; T, M.; S, M.; A, I.; A, T. Granular Tau Oligomers as Intermediates of Tau Filaments. *Biochemistry* **2007**, *46* (12), 3856–3861. https://doi.org/10.1021/BI061359O.

(170) S, M.; N, S.; Y, S.; S, M.; A, I.; A, T. Increased Levels of Granular Tau Oligomers: An Early Sign of Brain Aging and Alzheimer's Disease. *Neurosci. Res.* **2006**, *54* (3), 197–201. https://doi.org/10.1016/J.NEURES.2005.11.009.

(171) J, K.; EE, C.; G, L.; H, Y.; X, Y.; Q, Z. Evaluating Triggers and Enhancers of Tau Fibrillization. *Microsc. Res. Tech.* **2005**, *67* (3–4), 141–155. https://doi.org/10.1002/JEMT.20187.

(172) MS, F.; JQ, T.; VM, L. Neurodegenerative Diseases: A Decade of Discoveries Paves the Way for Therapeutic Breakthroughs. *Nat. Med.* **2004**, *10* (10), 1055–1063. https://doi.org/10.1038/NM1113.

(173) JQ, T.; MP, M. Overview of Protein Aggregation in Single, Double, and Triple

Neurodegenerative Brain Amyloidoses. *Neuromolecular Med.* **2003**, *4* (1–2), 1–5. https://doi.org/10.1385/NMM:4:1-2:1.

(174) TW, M.; J, N.; LY, H.; EJ, M.; JA, S.; EJ, C.; DA, B.; VM, L.; JQ, T.; SE, A. Novel Method to Quantify Neuropil Threads in Brains from Elders with or without Cognitive Impairment. *J. Histochem. Cytochem.* **2000**, *48* (12), 1627–1637. https://doi.org/10.1177/002215540004801206.

(175) JQ, T.; AB, S.; D, H.; VM, L. Microtubule-Stabilising Drugs for Therapy of Alzheimer's Disease and Other Neurodegenerative Disorders with Axonal Transport Impairments. *Expert Opin. Pharmacother.* **2005**, *6* (5), 683–686. https://doi.org/10.1517/14656566.6.5.683.

(176) S, R.; B, Z.; VM, L.; JQ, T. Axonal Transport Defects: A Common Theme in Neurodegenerative Diseases. *Acta Neuropathol.* **2005**, *109* (1), 5–13. https://doi.org/10.1007/S00401-004-0952-X.

(177) R, D.; JL, R.; YE, G.; EL, H. Differential Regulation of Dynein and Kinesin Motor Proteins by Tau. *Science* **2008**, *319* (5866), 1086–1089. https://doi.org/10.1126/SCIENCE.1152993.

(178) T, F.; J, E.; R, B. Tau-Mediated Cytotoxicity in a Pseudohyperphosphorylation Model of Alzheimer's Disease. *J. Neurosci.* **2002**, *22* (22), 9733–9741. https://doi.org/10.1523/JNEUROSCI.22-22-09733.2002.

(179) A, de C.; LM, F.; R, P.; GA, C.; BJ, B.; TL, S.-J.; BT, H. Caspase Activation Precedes and Leads to Tangles. *Nature* **2010**, *464* (7292), 1201–1204. https://doi.org/10.1038/NATURE08890.

(180) T, G.-I.; HL, W.; GW, R.; SD, H.; JH, G.; JJ, L.; TT, P.; LA, L.; BT, H. Clinical and Pathological Correlates of Apolipoprotein E Epsilon 4 in Alzheimer's Disease. *Ann. Neurol.* **1996**, *39* (1), 62–70. https://doi.org/10.1002/ANA.410390110.

(181) AB, R.; JL, C.; JM, A.; MS, K.; MA, T.-B.; J, L.; JI, L. Structural and Functional Changes

in Tau Mutant Mice Neurons Are Not Linked to the Presence of NFTs. *Exp. Neurol.* **2010**, *223* (2), 385–393. https://doi.org/10.1016/J.EXPNEUROL.2009.07.029.

(182) N, M.-C.; HG, Y.; V, K.; TL, S.-J.; D, C.; EA, S. Pathological Tau Disrupts Ongoing Network Activity. *Neuron* **2015**, *85* (5), 959–966. https://doi.org/10.1016/J.NEURON.2015.01.025.

(183) K, S.; J, L.; T, S.; J, P.; L, K.; M, I.; A, G.; M, D.; M, R.; E, M.; C, F.; M, Y.; J, O.; C, J.; A, M.; M, K.; B, H.; M, H.; KH, A. Tau Suppression in a Neurodegenerative Mouse Model Improves Memory Function. *Science* **2005**, *309* (5733), 476–481. https://doi.org/10.1126/SCIENCE.1113694.

(184) Zhao, X.; Kotilinek, L. A.; Smith, B.; Hlynialuk, C.; Zahs, K.; Ramsden, M.; Cleary, J.; Ashe, K. H. Caspase-2 Cleavage of Tau Reversibly Impairs Memory. *Nat. Med. 2016 2211* **2016**, *22* (11), 1268–1276. https://doi.org/10.1038/nm.4199.

(185) M, R.; L, K.; C, F.; J, P.; E, M.; K, S.; A, G.; M, Y.; J, L.; G, C.; M, H.; KH, A. Age-Dependent Neurofibrillary Tangle Formation, Neuron Loss, and Memory Impairment in a Mouse Model of Human Tauopathy (P301L). *J. Neurosci.* **2005**, *25* (46), 10637–10647. https://doi.org/10.1523/JNEUROSCI.3279-05.2005.

(186) ND, R.; M, W.; AJ, B.; A, B. MEROPS: The Database of Proteolytic Enzymes, Their Substrates and Inhibitors. *Nucleic Acids Res.* **2014**, *42* (Database issue). https://doi.org/10.1093/NAR/GKT953.

(187) Riedl, S. J.; Shi, Y. Molecular Mechanisms of Caspase Regulation during Apoptosis. *Nat. Rev. Mol. Cell Biol. 2004 511* **2004**, *5* (11), 897–907. https://doi.org/10.1038/nrm1496.

(188) T, O.; BT, F.; MM, S.; JW, L.; ND, W.; BC, R.; HE, P.; P, P.; MJ, R. Structural Analysis of Caspase-1 Inhibitors Derived from Tethering. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **2005**, *61* (Pt 5), 451–458. https://doi.org/10.1107/S1744309105010109.

(189) A, S.; C, B.; MG, G. Crystal Structure of Caspase-2, Apical Initiator of the Intrinsic Apoptotic Pathway. *J. Biol. Chem.* **2003**, *278* (43), 42441–42447.

https://doi.org/10.1074/JBC.M304895200.

(190) J, R.; DW, N.; KM, F.; M, G.; Y, G.; M, L.; EP, P.; DM, R.; R, R.; JP, V.; NA, T.; JW, B. The Three-Dimensional Structure of Apopain/CPP32, a Key Mediator of Apoptosis. *Nat. Struct. Biol.* **1996**, *3* (7), 619–625. https://doi.org/10.1038/NSB0796-619.

(191) I, M.; MB, L.; AJ, R.; C, D.; I, M.-S.; A, K. Structure of Human Caspase-6 in Complex with Z-VAD-FMK: New Peptide Binding Mode Observed for the Non-Canonical Caspase Conformation. *Bioorg. Med. Chem. Lett.* **2011**, *21* (18), 5244–5247. https://doi.org/10.1016/J.BMCL.2011.07.041.

(192) J, A.; B, F.; IT, W. Plasticity of S2-S4 Specificity Pockets of Executioner Caspase-7 Revealed by Structural and Kinetic Analysis. *FEBS J.* **2007**, *274* (18), 4752–4765. https://doi.org/10.1111/J.1742-4658.2007.05994.X.

(193) H, B.; M, D.; M, T.; L, K.; JC, W.; MG, G. Caspase-8 Specificity Probed at Subsite S(4): Crystal Structure of the Caspase-8-Z-DEVD-Cho Complex. *J. Mol. Biol.* **2000**, *302* (1), 9–16. https://doi.org/10.1006/JMBI.2000.4041.

(194) M, R.; HR, S.; FL, S.; RC, L.; GS, S. Dimer Formation Drives the Activation of the Cell Death Protease Caspase 9. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (25), 14250–14255. https://doi.org/10.1073/PNAS.231465798.

(195) Maillard, M. C.; Brookfield, F. A.; Courtney, S. M.; Eustache, F. M.; Gemkow, M. J.; Handel, R. K.; Johnson, L. C.; Johnson, P. D.; Kerry, M. A.; Krieger, F.; Meniconi, M.; Muñoz-Sanjuán, I.; Palfrey, J. J.; Park, H.; Schaertl, S.; Taylor, M. G.; Weddell, D.; Dominguez, C. Exploiting Differences in Caspase-2 and -3 S2 Subsites for Selectivity: Structure-Based Design, Solid-Phase Synthesis and in Vitro Activity of Novel Substrate-Based Caspase-2 Inhibitors. *Bioorg. Med. Chem.* **2011**, *19* (19), 5833–5851. https://doi.org/10.1016/J.BMC.2011.08.020.

(196) Thornberry, N. A.; Rano, T. A.; Peterson, E. P.; Rasper, D. M.; Timkey, T.; Garcia-Calvo, M.; Houtzager, V. M.; Nordstrom, P. A.; Roy, S.; Vaillancourt, J. P.; Chapman, K. T.;

Nicholson, D. W. A Combinatorial Approach Defines Specificities of Members of the Caspase Family and Granzyme B: FUNCTIONAL RELATIONSHIPS ESTABLISHED FOR KEY MEDIATORS OF APOPTOSIS. *J. Biol. Chem.* **1997**, *272* (29), 17907–17911. https://doi.org/10.1074/JBC.272.29.17907.

(197) Wang, Y. P.; Zhou, W.; Wang, J.; Huang, X.; Zuo, Y.; Wang, T. S.; Gao, X.; Xu, Y. Y.; Zou, S. W.; Liu, Y. Bin; Cheng, J. K.; Lei, Q. Y. Arginine Methylation of MDH1 by CARM1 Inhibits Glutamine Metabolism and Suppresses Pancreatic Cancer. *Mol. Cell* **2016**, *64* (4), 673–687. https://doi.org/10.1016/j.molcel.2016.09.028.

(198) Birktoft, J. J.; Rhodes, G.; Banaszak, L. J. Refined Crystal Structure of Cytoplasmic Malate Dehydrogenase at 2.5-Å Resolution. *Biochemistry* **1989**, *28* (14), 6065–6081. https://doi.org/10.1021/bi00440a051.

(199) Bergmeyer, H. U. *Methods of Enzymatic Analysis*, 2nd Editio.; 1974.

(200) Dragovich, P. S.; Fauber, B. P.; Corson, L. B.; Ding, C. Z.; Eigenbrot, C.; Ge, H.; Giannetti, A. M.; Hunsaker, T.; Labadie, S.; Liu, Y.; Malek, S.; Pan, B.; Peterson, D.; Pitts, K.; Purkey, H. E.; Sideris, S.; Ultsch, M.; VanderPorten, E.; Wei, B.; Xu, Q.; Yen, I.; Yue, Q.; Zhang, H.; Zhang, X. Identification of Substituted 2-Thio-6-Oxo-1,6-Dihydropyrimidines as Inhibitors of Human Lactate Dehydrogenase. *Bioorg. Med. Chem. Lett.* **2013**, *23* (11), 3186–3194. https://doi.org/10.1016/j.bmcl.2013.04.001.

(201) Vonrhein, C.; Flensburg, C.; Keller, P.; Sharff, A.; Smart, O.; Paciorek, W.; Womack, T.; Bricogne, G. Data Processing and Analysis with the AutoPROC Toolbox. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67* (4), 293–302. https://doi.org/10.1107/S0907444911007773.

(202) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40* (4), 658–674. https://doi.org/10.1107/S0021889807021206.

(203) PD, A.; PV, A.; G, B.; VB, C.; IW, D.; N, E.; JJ, H.; LW, H.; GJ, K.; RW, G.-K.; AJ, M.;

NW, M.; R, O.; RJ, R.; DC, R.; JS, R.; TC, T.; PH, Z. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66* (Pt 2), 213–221. https://doi.org/10.1107/S0907444909052925.

(204) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (4), 486–501. https://doi.org/10.1107/S0907444910007493.

(205) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. Overview of the CCP4 Suite and Current Developments. *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography April 18, 2011, pp 235–242. https://doi.org/10.1107/S0907444910045749.

(206) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (1), 12–21. https://doi.org/10.1107/S0907444909042073.

(207) Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkoczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L. W.; Jain, S.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D. Macromolecular Structure Determination Using X-Rays, Neutrons and Electrons: Recent Developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* **2019**, *75* (10), 861–877. https://doi.org/10.1107/S2059798319011471.

(208) Niesen, F. H.; Berglund, H.; Vedadi, M. The Use of Differential Scanning Fluorimetry to Detect Ligand Interactions That Promote Protein Stability. *Nat. Protoc.* **2007**, *2* (9), 2212–

2221. https://doi.org/10.1038/nprot.2007.321.

(209) Ericsson, U. B.; Hallberg, B. M.; DeTitta, G. T.; Dekker, N.; Nordlund, P. Thermofluor-Based High-Throughput Stability Optimization of Proteins for Structural Studies. *Anal. Biochem.* **2006**, *357* (2), 289–298. https://doi.org/10.1016/j.ab.2006.07.027.

(210) Laboratories, B.-R. *CFX Maestro™ Software for Mac User Guide*; 2017.

(211) Krissinel, E.; Henrick, K.; IUCr. Secondary-Structure Matching (SSM), a New Tool for Fast Protein Structure Alignment in Three Dimensions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (12), 2256–2268. https://doi.org/10.1107/S0907444904026460.

(212) I, B.; E, P. Biphasic Transition Curve on Denaturation of Chicken Cystatin by Guanidinium Chloride. Evidence for an Independently Unfolding Structural Region. *FEBS Lett.* **1992**, *299* (1), 66–68. https://doi.org/10.1016/0014-5793(92)80102-M.

(213) Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'Rule of Three' for Fragment-Based Lead Discovery? *Drug Discov. Today* **2003**, *8* (19), 876–877. https://doi.org/10.1016/S1359-6446(03)02831-9.

(214) Holdgate, G. A.; Ward, W. H. J. Measurements of Binding Thermodynamics in Drug Discovery. *Drug Discov. Today* **2005**, *10* (22), 1543–1550. https://doi.org/10.1016/S1359-6446(05)03610-X.

(215) PS, K.; AM, W.; MK, L.; SK, W.; J, O.; J, J.; C, S.; A, M.; N, B.; M, G. Large Scale Meta-Analysis of Fragment-Based Screening Campaigns: Privileged Fragments and Complementary Technologies. *J. Biomol. Screen.* **2015**, *20* (5), 588–596. https://doi.org/10.1177/1087057114565080.

(216) National Institute of Health - RePORTER https://reporter.nih.gov/search/w9d0MJZUxkGueAxJ85AJow/project-details/10071119 (accessed Aug 4, 2021).

(217) Denault, J.-B.; Salvesen, G. S. Expression, Purification, and Characterization of Caspases. *Curr. Protoc. Protein Sci.* **2002**, *30* (1), 21.13.1-21.13.15.

https://doi.org/10.1002/0471140864.PS2113S30.

(218) GraphPad Software. San Diego, California USA.

(219) Chauvier, D.; Renolleau, S.; Holifanjaniaina, S.; Ankri, S.; Bezault, M.; Schwendimann, L.; Rousset, C.; Casimir, R.; Hoebeke, J.; Smirnova, M.; Debret, G.; Trichet, A.-P.; Carlsson, Y.; Wang, X.; Bernard, E.; Hébert, M.; Rauzier, J.-M.; Matecki, S.; Lacampagne, A.; Rustin, P.; Mariani, J.; Hagberg, H.; Gressens, P.; Charriaut-Marlangue, C.; Jacotot, E. Targeting Neonatal Ischemic Brain Injury with a Pentapeptide-Based Irreversible Caspase Inhibitor. *Cell Death Dis.* **2011**, *2* (9), e203. https://doi.org/10.1038/CDDIS.2011.87.

(220) Boxer, M. B.; Quinn, A. M.; Shen, M.; Jadhav, A.; Leister, W.; Simeonov, A.; Auld, D. S.; Thomas, C. J. A Highly Potent and Selective Caspase 1 Inhibitor That Utilizes a Key 3-Cyanopropanoic Acid Moiety. *ChemMedChem* **2010**, *5* (5), 730–738. https://doi.org/10.1002/CMDC.200900531.

(221) Ingrid C. Choong, *; Willard Lew; Dennis Lee; Phuongly Pham; Matthew T. Burdett; Joni W. Lam; Christian Wiesmann; Tinh N. Luong; Bruce Fahr; Warren L. DeLano; Robert S. McDowell; Darin A. Allen; Daniel A. Erlanson; Eric M. Gordon, and; O'Brien, T. Identification of Potent and Selective Small-Molecule Inhibitors of Caspase-3 through the Use of Extended Tethering and Structure-Based Drug Design. *J. Med. Chem.* **2002**, *45* (23), 5005–5022. https://doi.org/10.1021/JM020230J.

(222) Walters, J.; Schipper, J. L.; Swartz, P.; Mattos, C.; Clark, A. C. Allosteric Modulation of Caspase 3 through Mutagenesis. *Biosci. Rep.* **2012**, *32* (Pt 4), 401. https://doi.org/10.1042/BSR20120037.

(223) Fang, B.; Guoxing, A. E.; Ae, F.; Agniswamy, J.; Robert, A. E.; Ae, W. H.; Weber, I. T. Caspase-3 Binds Diverse P4 Residues in Peptides as Revealed by Crystallography and Structural Modeling. https://doi.org/10.1007/s10495-009-0333-y.

(224) B, F.; PI, B.; J, T.; IT, W. Structural and Kinetic Analysis of Caspase-3 Reveals Role for

S5 Binding Site in Substrate Recognition. *J. Mol. Biol.* **2006**, *360* (3), 654–666. https://doi.org/10.1016/J.JMB.2006.05.041.

(225) Stennicke, H. R.; Salvesen, G. S. Biochemical Characteristics of Caspases-3, -6, -7, and -8. *J. Biol. Chem.* **1997**, *272* (41), 25719–25723. https://doi.org/10.1074/JBC.272.41.25719.

(226) Wang, J. Estimation of the Quality of Refined Protein Crystal Structures. *Protein Sci.* **2015**, *24* (5), 661. https://doi.org/10.1002/PRO.2639.

(227) Liu, L.-K. *Discovery of Small Molecule Inhibitors of Hyaluronan Binding at Cell Receptor CD44* ; 2015.

(228) Ichihara, O.; Barker, J.; Law, R. J.; Whittaker, M. Compound Design by Fragment-Linking. *Mol. Inform.* **2011**, *30* (4), 298–306. https://doi.org/10.1002/MINF.201000174.

(229) Jencks, W. P. On the Attribution and Additivity of Binding Energies. *Proc. Natl. Acad. Sci.* **1981**, *78* (7), 4046–4050. https://doi.org/10.1073/PNAS.78.7.4046.

(230) Murray, C. W.; Verdonk, M. L. Entropic Consequences of Linking Ligands. *Fragm. Approaches Drug Discov.* **2006**, *34*, 55–66. https://doi.org/10.1002/3527608761.CH3.

(231) Röhrig, C. H.; Loch, C.; Guan, J.-Y.; Siegal, G.; Overhand, M. Fragment-Based Synthesis and SAR of Modified FKBP Ligands: Influence of Different Linking on Binding Affinity. *ChemMedChem* **2007**, *2* (7), 1054–1070. https://doi.org/10.1002/CMDC.200600296.

(232) Borsi, V.; Calderone, V.; Fragai, M.; Luchinat, C.; Sarti, N. Entropic Contribution to the Linking Coefficient in Fragment Based Drug Design: A Case Study. *J. Med. Chem.* **2010**, *53* (10), 4285–4289. https://doi.org/10.1021/jm901723z.

(233) Maestro. Schrodinger: New York, NY 2021.

(234) Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Mol. Des. 2013 273* **2013**, *27* (3), 221–234. https://doi.org/10.1007/S10822-013-9644-8.

(235) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the Role of the Crystal Environment in Determining Protein Side-Chain Conformations. *J. Mol. Biol.* **2002**, *320* (3), 597–608. https://doi.org/10.1016/S0022-2836(02)00470-9.

(236) LigPrep. Schrodinger: New York, NY 2021.

(237) Lebedev, A. A.; Young, P.; Isupov, M. N.; Moroz, O. V; Vagin, A. A.; Murshudov, G. N. JLigand: A Graphical Tool for the CCP4 Template-Restraint Library. *Acta Crystallogr. D. Biol. Crystallogr.* **2012**, *68* (Pt 4), 431–440. https://doi.org/10.1107/S090744491200251X.

(238) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2015**, *12* (1), 281–296. https://doi.org/10.1021/ACS.JCTC.5B00864.

(239) Greenwood, J. R.; Calkins, D.; Sullivan, A. P.; Shelley, J. C. Towards the Comprehensive, Rapid, and Accurate Prediction of the Favorable Tautomeric States of Drug-like Molecules in Aqueous Solution. *J. Comput. Mol. Des. 2010 246* **2010**, *24* (6), 591–604. https://doi.org/10.1007/S10822-010-9349-1.

(240) Richard A. Friesner, *; Robert B. Murphy, †; Matthew P. Repasky, †; Leah L. Frye, ‡; Jeremy R. Greenwood, †; Thomas A. Halgren, †; Paul C. Sanschagrin, † and; Mainz†, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *J. Med. Chem.* **2006**, *49* (21), 6177–6196. https://doi.org/10.1021/JM051256O.

(241) Watts, K. S.; Dalal, P.; Murphy, R. B.; Sherman, W.; Friesner, R. A.; Shelley, J. C. ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers. *J. Chem. Inf. Model.* **2010**, *50* (4), 534–546. https://doi.org/10.1021/CI100015J.

(242) Mahoney, D. J.; Aplin, R. T.; Calabro, A.; Hascall, V. C.; Day, A. J. Novel Methods for

the Preparation and Characterization of Hyaluronan Oligosaccharides of Defined Length. *Glycobiology* **2001**, *11* (12), 1025–1033. https://doi.org/10.1093/glycob/11.12.1025.

(243) Emsley, P.; Cowtan, K. Coot : Model-Building Tools for Molecular Graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (12), 2126–2132. https://doi.org/10.1107/S0907444904019158.

(244) Vanommeslaeghe, K.; Guvench, O.; MacKerell, A. D.; Jr. Molecular Mechanics. *Curr. Pharm. Des.* **2014**, *20* (20), 3281.

(245) Hospital, A.; Goñi, J. R.; Orozco, M.; Gelpí, J. L. Molecular Dynamics Simulations: Advances and Applications. *Adv. Appl. Bioinform. Chem.* **2015**, *8* (1), 37. https://doi.org/10.2147/AABC.S70333.

(246) Di Carlo, D. T.; Cagnazzo, F.; Benedetto, N.; Morganti, R.; Perrini, P. Multiple High-Grade Gliomas: Epidemiology, Management, and Outcome. A Systematic Review and Meta-Analysis. *Neurosurgical Review*. Springer Verlag June 1, 2019, pp 263–275. https://doi.org/10.1007/s10143-017-0928-7.

(247) Gallego, O. Nonsurgical Treatment of Recurrent Glioblastoma. *Current Oncology*. Multimed Inc. August 20, 2015, pp e273–e281. https://doi.org/10.3747/co.22.2436.

(248) McGuire, S. World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015. *Advances in Nutrition*. American Society for Nutrition March 1, 2016, pp 418–419. https://doi.org/10.3945/an.116.012211.

(249) Anido, J.; Sáez-Borderías, A.; Gonzàlez-Juncà, A.; Rodón, L.; Folch, G.; Carmona, M. A.; Prieto-Sánchez, R. M.; Barba, I.; Martínez-Sáez, E.; Prudkin, L.; Cuartas, I.; Raventós, C.; Martínez-Ricarte, F.; Poca, M. A.; García-Dorado, D.; Lahn, M. M.; Yingling, J. M.; Rodón, J.; Sahuquillo, J.; Baselga, J.; Seoane, J. TGF-β Receptor Inhibitors Target the CD44high/Id1high Glioma-Initiating Cell Population in Human Glioblastoma. *Cancer Cell* **2010**, *18* (6), 655–668. https://doi.org/10.1016/j.ccr.2010.10.023.

(250) Lesley, J.; Hyman, R.; Kincade, P. W. CD44 and Its Interaction with Extracellular Matrix. *Adv. Immunol.* **1993**, *54* (C), 271–335. https://doi.org/10.1016/S0065-2776(08)60537-4.

(251) Bennett, K. L.; Jackson, D. G.; Simon, J. C.; Tanczos, E.; Peach, R.; Modrell, B.; Stamenkovic, I.; Plowman, G.; Aruffo, A. Cd44 Isoforms Containing Exon V3 Are Responsible for the Presentation of Heparin-Binding Growth Factor. *J. Cell Biol.* **1995**, *128* (4), 687–698. https://doi.org/10.1083/jcb.128.4.687.

(252) Pietras, A.; Katz, A. M.; Ekström, E. J.; Wee, B.; Halliday, J. J.; Pitter, K. L.; Werbeck, J. L.; Amankulor, N. M.; Huse, J. T.; Holland, E. C. Osteopontin-CD44 Signaling in the Glioma Perivascular Niche Enhances Cancer Stem Cell Phenotypes and Promotes Aggressive Tumor Growth. *Cell Stem Cell* **2014**, *14* (3), 357–369. https://doi.org/10.1016/j.stem.2014.01.005.

(253) Barat, S.; Chen, X.; Bui, K. C.; Bozko, P.; Götze, J.; Christgen, M.; Krech, T.; Malek, N. P.; Plentz, R. R. Gamma-Secretase Inhibitor IX (GSI) Impairs Concomitant Activation of Notch and Wnt-Beta-Catenin Pathways in CD44+ Gastric Cancer Stem Cells. *Stem Cells Transl. Med.* **2017**, *6* (3), 819–829. https://doi.org/10.1002/SCTM.16-0335.

(254) Hartmann, M.; Parra, L. M.; Ruschel, A.; Lindner, C.; Morrison, H.; Herrlich, A.; Herrlich, P. Inside-out Regulation of Ectodomain Cleavage of Cluster-of-Differentiation-44 (CD44) and of Neuregulin-1 Requires Substrate Dimerization. *J. Biol. Chem.* **2015**, *290* (28), 17041–17054. https://doi.org/10.1074/jbc.M114.610204.

(255) Georgiev, M.; Pastore, S.; Lulli, D.; Alipieva, K.; Kostyuk, V.; Potapovich, A.; Panetta, M.; Korkina, L. Verbascum Xanthophoeniceum-Derived Phenylethanoid Glycosides Are Potent Inhibitors of Inflammatory Chemokines in Dormant and Interferon-Gamma-Stimulated Human Keratinocytes. *J. Ethnopharmacol.* **2012**, *144* (3), 754–760. https://doi.org/10.1016/j.jep.2012.10.035.

(256) Li, L.; Tsao, R.; Liu, Z.; Liu, S.; Yang, R.; Young, J. C.; Zhu, H.; Deng, Z.; Xie, M.; Fu, Z. Isolation and Purification of Acteoside and Isoacteoside from Plantago Psyllium L. by

High-Speed Counter-Current Chromatography. *J. Chromatogr. A* **2005**, *1063* (1–2), 161–169. https://doi.org/10.1016/j.chroma.2004.11.024.

(257) Akdemir, Z.; Kahraman, Ç.; Tatli, I. I.; Küpeli Akkol, E.; Süntar, I.; Keles, H. Bioassay-Guided Isolation of Anti-Inflammatory, Antinociceptive and Wound Healer Glycosides from the Flowers of Verbascum Mucronatum Lam. *J. Ethnopharmacol.* **2011**, *136* (3), 436–443. https://doi.org/10.1016/j.jep.2010.05.059.

(258) V, S.; IR, V.; H, G.; E, C.; H, D.; S, J. CD44 Regulates Cell Migration in Human Colon Cancer Cells via Lyn Kinase and AKT Phosphorylation. *Exp. Mol. Pathol.* **2007**, *83* (2), 207–215. https://doi.org/10.1016/J.YEXMP.2007.04.008.

(259) K, N.; S, O.; KM, L.; SA, Y.; I, S. CD44 Regulates Cell Proliferation, Migration, and Invasion via Modulation of c-Src Transcription in Human Breast Cancer Cells. *Cell. Signal.* **2015**, *27* (9), 1882–1894. https://doi.org/10.1016/J.CELLSIG.2015.05.002.

(260) Hsin, K. Y.; Matsuoka, Y.; Asai, Y.; Kamiyoshi, K.; Watanabe, T.; Kawaoka, Y.; Kitano, H. SystemsDock: A Web Server for Network Pharmacology-Based Prediction and Analysis. *Nucleic Acids Res.* **2016**, *44* (W1), W507–W513. https://doi.org/10.1093/nar/gkw335.

(261) Krasavin, M.; Kalinin, S.; Zozulya, S.; Gryniukova, A.; Borysko, P.; Angeli, A.; Supuran, C. T. Screening of Benzenesulfonamide in Combination with Chemically Diverse Fragments against Carbonic Anhydrase by Differential Scanning Fluorimetry. *https://doi.org/10.1080/14756366.2019.1698562* **2019**, *35* (1), 306–310. https://doi.org/10.1080/14756366.2019.1698562.

(262) Hamiaux, C.; Drummond, R. S. M.; Janssen, B. J.; Ledger, S. E.; Cooney, J. M.; Newcomb, R. D.; Snowden, K. C. DAD2 Is an α/β Hydrolase Likely to Be Involved in the Perception of the Plant Branching Hormone, Strigolactone. *Curr. Biol.* **2012**, *22* (21), 2032–2036. https://doi.org/10.1016/J.CUB.2012.08.007.

(263) Pacold, M. E.; Brimacombe, K. R.; Chan, S. H.; Rohde, J. M.; Lewis, C. A.; Swier, L. J.

Y. M.; Possemato, R.; Chen, W. W.; Sullivan, L. B.; Fiske, B. P.; Cho, S.; Freinkman, E.; Birsoy, K.; Abu-Remaileh, M.; Shaul, Y. D.; Liu, C. M.; Zhou, M.; Koh, M. J.; Chung, H.; Davidson, S. M.; Luengo, A.; Wang, A. Q.; Xu, X.; Yasgar, A.; Liu, L.; Rai, G.; Westover, K. D.; Heiden, M. G. Vander; Shen, M.; Gray, N. S.; Boxer, M. B.; Sabatini, D. M. A PHGDH Inhibitor Reveals Coordination of Serine Synthesis and One-Carbon Unit Fate. *Nat. Chem. Biol. 2016 126* **2016**, *12* (6), 452–458. https://doi.org/10.1038/nchembio.2070.

(264) Vedadi, M.; Niesen, F. H.; Allali-Hassani, A.; Fedorov, O. Y.; Finerty, P. J.; Wasney, G. A.; Yeung, R.; Arrowsmith, C.; Ball, L. J.; Berglund, H.; Hui, R.; Marsden, B. D.; Nordlund, P.; Sundstrom, M.; Weigelt, J.; Edwards, A. M. Chemical Screening Methods to Identify Ligands That Promote Protein Stability, Protein Crystallization, and Structure Determination. *Proc. Natl. Acad. Sci.* **2006**, *103* (43), 15835–15840. https://doi.org/10.1073/PNAS.0605224103.

(265) Senisterra, G. A.; Markin, E.; Yamazaki, K.; Hui, R.; Vedadi, M.; Awrey, D. E. Screening for Ligands Using a Generic and High-Throughput Light-Scattering-Based Assay: *http://dx.doi.org/10.1177/1087057106294699* **2016**, *11* (8), 940–948. https://doi.org/10.1177/1087057106294699.

(266) Daumantas Matulis, ‡; James K. Kranz; F. Raymond Salemme, and; Todd*, M. J. Thermodynamic Stability of Carbonic Anhydrase: Measurements of Binding Affinity and Stoichiometry Using ThermoFluor. *Biochemistry* **2005**, *44* (13), 5258–5266. https://doi.org/10.1021/BI048135V.

(267) Sparks, R. P.; Jenkins, J. L.; Fratti, R. Use of Surface Plasmon Resonance (SPR) to Determine Binding Affinities and Kinetic Parameters Between Components Important in Fusion Machinery. *Methods Mol. Biol.* **2019**, *1860*, 199–210. https://doi.org/10.1007/978-1-4939-8760-3_12.

(268) Biacore Assay Handbook.

(269) Toole, B. P. Hyaluronan-CD44 Interactions in Cancer: Paradoxes and Possibilities. *Clin. Cancer Res.* **2009**, *15* (24), 7462–7468. https://doi.org/10.1158/1078-0432.CCR-09-0479.

(270) NOBLE, P. W.; LIANG, J.; JIANG, D. Hyaluronan as an Immune Regulator in Human Diseases. *Physiol. Rev.* **2011**, *91* (1), 221. https://doi.org/10.1152/PHYSREV.00052.2009.

(271) MS, P.; PH, W. Hyaluronic Acid Receptor for Endocytosis (HARE)-Mediated Endocytosis of Hyaluronan, Heparin, Dermatan Sulfate, and Acetylated Low Density Lipoprotein (AcLDL), but Not Chondroitin Sulfate Types A, C, D, or E, Activates NF-KB-Regulated Gene Expression. *J. Biol. Chem.* **2014**, *289* (3), 1756–1767. https://doi.org/10.1074/JBC.M113.510339.

(272) MG, S.; L, D.; LB, T.; GD, G.; Y, Z.; BP, T. Inhibition of Functional Hyaluronan-CD44 Interactions in CD133-Positive Primary Human Ovarian Carcinoma Cells by Small Hyaluronan Oligosaccharides. *Clin. Cancer Res.* **2009**, *15* (24), 7593–7601. https://doi.org/10.1158/1078-0432.CCR-09-2317.

(273) AG, G.; SL, T.; LB, T.; WG, W.; RP, V.; JD, D.; FV, K.; LN, B.; BP, T.; BL, M. Targeting Hyaluronan Interactions in Malignant Gliomas and Their Drug-Resistant Multipotent Progenitors. *Clin. Cancer Res.* **2008**, *14* (6), 1804–1813. https://doi.org/10.1158/1078-0432.CCR-07-1228.

(274) Liu, K.; Huang, X. *Ongoing Work - No Title Yet*; 2021.

(275) Ugi, I. Versuche Mit Isonitrilen. *Angew. Chem. Int. Ed. Engl.* **1959**, *71*, 386–386.